NWX-US DEPT OF COMMERCE

Demystifying the Census API
July 22, 2020
1:00 pm CT

Coordinator: 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 a 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 your host Kim Davis. Thank you. You may begin.

Kim Davis: Thank you. Good day everyone. Thank you for joining us for another Census Academy webinar today. Today's presentation will be Demystifying the Census API. We are recording the webinar. It will be available on Census Academy within the next couple of weeks for your reference and review. We have a large number of participants on the webinar today. So if we don't get through all of the questions and answers at the end of today's session or in the chat, we will provide follow-up contact information for you at the end of the presentation and we will follow up with those that we have contact information.

We will not be able to address any of the 2020 Census operations questions that you may have today. You're certainly welcome to go to 2020census.gov for more details about the 2020 operation. And co-hosting with me today and moderating the chat will be Deb Rivera Nieves. Our facilitator today will be Jeana. And our guest speaker today is Mr. Tyson Weister. He's a survey statistician at the US Census Bureau where he engages users in the future of accessing Census data, and provides training on data.census.gov.

As part of the Census Bureau's centralized Data Dissemination Program, he also uses - excuse me - helps users access Census data API and Micro Data access tools. Welcome, Tyson.

Tyson Weister: Great. Thanks, Kim and thank you all for joining the webinar this afternoon on Demystifying the Census API. Whether you are brand new to using an API in general or if you're familiar with APIs but not familiar with the layout of the Census Bureau API, this webinar will serve as an introduction to help get you started and get you familiar with how to navigate the application programming interface.

If you're new to it, it may sound like a lot at first but it's really as simple as finding a sample URL, editing that URL to get the data that you're looking for. And before we get there we're going to show some basic information about the API. After we do that will look at a sample API URL and show you the different components of it just to give you a little bit of familiarity with the URLs you'll be working with. And then the majority of today we'll focus on a

live demonstration of the API in action, where we will go to the site. We'll work through three quick examples and show you how you can view the results on your screen and how you can save the information as a CSV file that you can open up and reformat in Excel. And we'll wrap up with some resources for learning more and of course question and answer.

From the Census Bureau's application programming interface, there are three APIs that we have at the Census Bureau. They're all free, publicly accessible and open source services to get our data. There is the Census data API that has the raw statistical data from the surveys and programs across the Census Bureau. There's also a Geocoder portion of the API where you can search by a specific address and get a result of the Census geographic areas that that address falls within. And then for geographers as well there's the TIGERweb Services that has the boundaries and shapes for mapping.

Most folks using the API are interested in the data API, and that's what we're going to focus on today. The application programming interface is the standardized way for users and software developers to access the public data. It supports mobile and web applications, drives interactive data visualization, and the way that this happens is because different programs can connect directly to the API and request that data in real-time. We'll take a look at an example of that in just a moment - I've something that we do internally - but because this is a public API, anyone on the external side can also create their own web applications and visualizations from our API as well and connect it in real-time.

The advantage of the API is it lets you find the exact variables and geographies that you're looking for. There is immediate access to the updates. Rather than having to go through the process, download the data. If there were any updates to the underlying data, because the API is getting this information in real-time and feeding it over to you, you would see those updates. And there's no need to host the data or download it and store it on your own server.

And then finally, the last advantage of the API is it has more data than what you can find on data.census.gov. So even if you're not a programmer or developer, you may find the API easier to work with to find some of the information you're looking for rather than perhaps a set of downloadable files from the different surveys and programs for the FTP site. We're going to walk through some basic examples showing you how to get the exact data of your looking for. And as an example of accessing data that's not available on data.census.gov, we'll show some of the examples looking at data from before the 2010 American Community Survey - which isn't currently on data.census.gov.

The API for different web applications and programs can be used to power what you're saying on the screen. So for the Census Bureau's primary dissemination platform data.census.gov, the API is at the heart of everything that we're doing. It is feeding all of the charts, tables, and maps that you see when you're on data.census.gov, as you navigate through that site. On the back end, calls are being made to the API and feeding all that information to you onscreen in a very user-friendly format.

Today we're going to learn how to access that information directly from the API. There are 488 different data sets on the API that fit into the 20 different high-level surveys and programs. So just to give you a quick glance as to what's available, we have information from demographic and economic programs as well as the planning database. Today we're going to walk through some examples from the American Community Survey and the Population Estimates Program.

The available geography for the different data sets depends on the data set for that survey and program and what levels of geography they produce the information for. So in general, you could get data for all of these different geographies you see on the screen through our API as long as that survey provides the data at that level of geography. If you're new to Census data, please be aware that the decennial census provides the most geographic coverage, followed by the American Community Survey. And then here on the screen, the geographic area is highlighted in red have data through our economic surveys and programs like the economic census.

In order to access the different options of these geography levels in the 488 datasets, where you want to go is the developer's page as your first stop for the public API. We're on census.gov/developers. It will take you here and we'll go through this in just a moment, but I just want to point out at the very top of the page will be a link to the discovery tool. That is the most up-to-date listing of all the available data sets on our site.

Also on the developer page are guidance for things like the developer's forum,

where you can ask questions to other users like yourself, the API user guide, and then on the left-hand side is an option to request a key. An API key is free and it's only needed if you're making more than 500 API calls per day. So we're not going to get into the API key for this demonstration but where we will go is to the discovery tool to start digging into some examples from the data sets that are available.

Once you click into the discovery tool, we'll open up the HTML version so we can see the information on screen. And for each of the data sets, you will be able to click into the available geography, variables, examples, as well as some additional links for documentation and group calls. We'll check this out in just a moment on the live site.

But before we get there, just want to get some familiarity with how to build an API request and what the different portions of the URL represent. You certainly don't need to memorize all the different portions but it is helpful to be familiar with the general scope of how they're laid out and how they're separated. So every API call for the data API will begin with API census.gov/data.

Next, you will choose one of the 488 data sets. In this API call example, we're working out of the 2018 vintage of the Population Estimates Program and we're looking at the characteristics of each group. After your data set will be question mark get equals. The fourth part will be where you add your variables. And this particular example, we added a variable for pop to get the population total, geo name so we can see the label for our geography. Date

description and date code equals 11, which gives us the data as of July 1st, 2018 and race equals ten. This will restrict our results for the Asian alone or in combination population. For each data set when you click the variable section of the discovery tool you'll be able to see all of the available options.

And then the last part is the geography portion of the URL. Here we're looking at ampersand for county, then you see that asterisk or wild card, ampersand in state 24. Anywhere you see a wild card gives you results for all so this is telling us we're getting results for all counties in state 24 which is the two-digit code for Maryland. Once you put that all together, again we won't be doing this from scratch. We'll work off of the example URL that's provided and make small tweaks to it in order to get what we're looking for.

And when you open that result, you will be able to see the data in your web browser as long as you're using Google Chrome is what we're going to use today. And this is what the data look like. Each geography has its own line and you could see the results. So as an example, for Allegany County, Maryland we can see the population of 1,017 which represents the population for the Asian alone or in combination population as of July 1st, 2018.

So with that basic information let's go ahead and transition to the live demonstration here. With a demo, our goal is for you all to walk out feeling comfortable on how to find the available data sets in the API, how to find that sample URL and navigate the discovery tool to edit from the sample URL to the geographies that you're specifically interested in, and the variables you're interested in to access the data you need. You will also be able to leave and

see the results in your web browser as well as save them as a CSV file and format it into a readable format.

In order to get there, we're going to work through three examples. The first example is going to be very basic. For a single estimate in geography, what was the Hispanic population in California as of July 1st, 2019? So we'll go ahead and navigate here. I have Google Chrome open and I'm already on census.gov/developer. You all are welcome to follow along or if you just want to watch, I do have step-by-step PowerPoint slides that are provided with this, that walk through everything. So if you missed a step here or there, know that you have this to look back to with the slides as well as the recording for this particular webinar.

On the developer page, there are a couple different ways you can go about navigating. The first way I'm going to show you is the available API just because it's a little bit less overwhelming to navigate through as a starting point if you're just new. And then you get these high-level surveys and programs that I have listed in one of those files earlier are laid out here. So there is a page of results. I want data from the Population Estimates Program for the Hispanic population. And it's all alphabetical. So clicking over to page two, I get to the Ps and I see population estimates and projections.

Clicking on population estimates and projections, here it prompts us for a couple of different options whether we want population estimates, projections, international databases, as well as some additional related information. I want data for the population estimates. It's the official source of population totals in

between the decennial census years for the geographies that are covered.

And here we get the page for population estimates API. We can see the different vintages laid out here for us as well as different categories of data sets to choose from. So population estimates if I just wanted totals, components of change for things like births, deaths, migrations, monthly population estimates.

And here is where I'm getting at what I'm interested in - demographic characteristic estimates by age group - and housing count information. When you find what you're looking for here under demographic characteristic estimates by age group, I can confirm that what I'm interested in is a selectable variable. I always like to click on the examples in support of geography. I'm going to right-click that link and open it in a new tab.

 And this takes us to the discovery tool. But it's filtered for the one exact data set that we're interested in rather than all 488 data sets that are available. When I'm working through this, I like to work out of the examples and the variable sections. The examples give you that sample URL to work off of as well as a listing of all of the different geographic levels that are available. And then the variables allow you to specify what date are you want to include in your call.

So right-click examples and open it in a new tab. And this particular data set is provided for just a few different geographic levels. You have the United States, the state, and counties. We want Hispanic population for California so

I'm going to be working out of the state section. And I can see within the state - which is summary level 040 - there are two example URLs that I can choose from. So there is ampersand for state wildcard as one example URL. That gives me data for all states in the United States. And the second example ampersand for state home 01 which will give me data for Alabama.

I'm going to choose this first example here, copying that URL to get data for all states. That way I can find what the state code is for California. And I'm opening that in a new tab here. Paste that. And you want to make sure to delete the portion of the URL that says ampersand key equals your key goes here, and press enter. And just as we work through this here, one thing I'm going to do so you all can see just a little bit better, increase the size of the URL for you.

So this is what we're going to be working out as we edit the geography in the URL, running that gave us data for all states. And from here I can do control f and start typing in California and see that we have the data for the total population and what we want state code 06. In our URL to get data for California you just replace the wild card with 06 and press enter. And I like to press enter in between each step as I edit the URL in case I make a minor typo. I'll know exactly where I went wrong when I'm going through this.

Now we have data for total population in California but we just need to replace the variable in the sample call, with the variable for Hispanic. So navigating backward in our navigation, and this time instead of clicking into example we'll click on variable. And we can see the different variables that are

available in this data set on the left-hand side as well as the label. So they're pretty easy to read through in this particular data set. You can see right off the bat for HISP provides data by Hispanic origin.

Anywhere that you see default displayed in the required column, that means it's a categorical variable and if you want to see what the values represent, one thing that we can do in a lot of data sets is just click on the variable name on the left. So I'm going to click on HISP. These categorical variables aren't particularly common in the American Community Survey or the Decennial Census but they are used pretty regularly in many other data sets on our API. So you can see HISP value of two gives you data for the Hispanic population, a value of one for non-Hispanic, and a value of zero for the total population.

So that's the variable we want HISP. So navigating back to our tab that we've been modifying, for that sample call you just want to click into the variable section of your URL and begin adding to it. I'm going to click and put a comma and then type in HISP and press enter. And we can see our results have been updated. So we have the total population for those groups with code zero, non-Hispanic population of code one, and the Hispanic population data for code two.

At this point, we could go ahead and use the results from here but if we wanted to modify that call to just give us data for that group code two for Hispanic, you can click into the URL and add ampersand HISP equals two, and press enter.

And that's how simple it is just to edit that URL to get the data. We made a couple of small tweaks from getting data for total population in all states to narrow that down to California and then to provide data for the Hispanic population.

Moving on to our next example, before we dig into another data set I just want to give a little bit of background information. But what we're going to show next is just how you can do something very similar to add more variables and more geographies into the same call, in case you were interested in more data. In the example, we'll work through the American Community Survey and look at median income and homeownership rates across all counties in California. And we're going to look at that 2007 data as one example of a year that isn't available on data.census.gov.

Before we get in there just as a note, the American Community Survey and Decennial Census provide so much data. And because of that, their API uses code-based variable names. So in the population estimates when we clicked variables there were 18 different variables in that data set, which makes it easy for the variables to be labeled in terms of something that makes sense like HISP for Hispanic.

For the American Community Survey, as an example, one of their data sets has 35,000 different variables. So they have codes in place to represent what those variables mean. It may look random at first but we'll explain it to you on the next slide. Right now we're just showing an example of a variable from the American Community Survey B17001_002E. And that gives you data for the

total people in poverty.

Although it looks random, that variable name has meaning. It always starts with the table ID. B17001 is the table ID that we're showing here from data.census.gov for poverty status in the past 12 months by sex by age. The next part of the variable 002 refers to the line of the table. So this is getting us data from the second row that has data in this table - income in the past 12 months below the poverty level. Variables that end in E give you data for the estimate and variables that end in M give you data for the margin of error.

As you're working through the American Community Survey, you'll also see variables that end in PE which gives you the percent estimate and PM for the percent margin of error.

And as a final note before we dive into the ACS data sets on the API, just wanted to let you know that you do have to choose the type of table that you're interested in because when you're working through the API, that's how the data sets are organized in that particular tool. So all of your calls for American Community Survey data we'll start with one of these where you have to choose ACS1 to get the detail table, profile to get the data profiles, and so on.

If you're not sure where to start, we recommended data profiles because they cover all topics of the American Community Survey in just a few tables, which translates into less variables the search through on the API. They're the most popular collection of ACS statistics and they include estimate as well as

the percentages. So for our example, something high-level like median income in homeownership will definitely be included in those data profiles.

So navigating here just to close out some of what we were working on earlier. On the census.gov/developers page, I showed you one way you can go about navigating which is through available APIs. This time we're going to dig right into the discovery tool to show you the alternative pathways to navigate. When you click discovery tool, you always want to use dot HTML generally if you're a new user. That way you can view everything in your web browser.

And as we zoom out, you'll notice this looks really similar to where we were at earlier. But it loads all 488 options in one view as opposed to making a series of clicks that gets you to a tailored view of the discovery tool. When you're on this page though, it gives you the most up-to-date options that are available to use. So something like the old American Community Survey data that was recently added, this is the best pathway to access that information. And you can also use control F on this page and type in a keyword. Something like 2007 will get you to the data that were available.

And the American Community Survey is the first data set that's available. I can see here American Community Survey detail tables are the first option in the second option, American Community Survey one-year estimate data profiles, which is the date of profiles is where you want to start for that high-level ACS information.

I'm going to go ahead and open up the example in a new tab. And while I'm

on the screen I'm going to go ahead and open up the variables as well. And then in my tab here - the first tab and I have open - will be for the example calls. So like before we see the list of available geographies on the left. We know this in comparison to the American Community Survey - even the one-year estimates - cover much more geographies than the population estimates. We wanted data at the county level so I see here state, county, summary level 050. And there are three different examples to choose from.

Looking at the tail end you can see how they differ. The first example provides us data ampersand for county wildcard to get data for all counties. The second example gives us data for all counties in all states. This really gives us the same information as the first example but it's just a different way of writing it out. And then the last example ampersand for county 003 and state code 1 would give us data for Baldwin County, Alabama.

I want data for all counties in California so I'm going to happen to choose the second example as my starting point, copy that link address, and paste it into a new tab and then press enter. So here we get data for all counties in the United States. If I need that reminder I can start type in California. In reference, the state code is 06.

So going back to our URL just like before we can replace the wild card next to state to 06 and press enter. And notice that our results are reflected to give us data for all counties in California. Because we're in the one-year estimates, it's giving us data for counties that have populations of 65,000 or more. So we have our geography of interest now and data and year. We just need to replace

the sample URL variable with the variables that we're interested in for median income and homeownership.

And I had opened up that variables page earlier here on another tab and it loaded the available options to us. Noticed that there's much more than we saw on the population estimate. However, you can use control F and type in a keyword when you're working in the data profiles to get where you want to go pretty easily. So I'm going to press control F and type in median and it actually takes me to the variable I'm interested in first - DP03_0063E provides data for the estimate in the income section of data profile 3. And out of total households, it gives us the median household income which is the data that we're looking for.

Copying that variable DP03_0063E we can navigate back to the tab that we're working off of from our API URL and replace the sample variable with the variable we're interested in for median income and press enter. Now we have results for median income across all counties in California.

However, we also wanted this data for homeownership. So we can go ahead and navigate back to that variables page. I apologize here. Actually going to close some of this up that I'm not interested in. Okay. So we did the control F for median. Now we'll do it for owner. And you can see the first result that it gives us, the homeowner vacancy rate - which is not what am interested in. I want to know how many housing units are occupied by the owner.

So as we enter and page through the different results, we can see the next set

of results for DP04_0045E gives us the estimate for the total number of owner-occupied housing units. And to get the rate we would want to use the percent estimate for occupied housing units that are occupied by the owner versus the renter. So we can see by carefully reading the labels that DP04_0045PE gives us the percent estimate for homeownership.

So navigating back to our tab where we have the median income here in our variable name, we will just edit the URL. And next to DP03_0063E you can press comma and then paste the next variable of interest and press enter. You can add up to 50 variables in a single call. And the nice thing about this is in comparison to using data.census.gov or other access points to get data in bulk, this allows you to specify the exact variables you're interested in rather than having to download an entire table or an entire data set and then deleting all of the information that you're not interested in. This allows you to specify and get only what you're interested in from the beginning.

Here in our results, we can see for Alameda County, California the median household income was 68,740 and the homeownership rate in 2007 was 57.4%.

Now, the other thing that you may want to work with on the API is how do I get results for an entire table. There is group functionality specifically that allows you to get the results for an entire table without having to specify all of the variables individually and without regard to whether that table may have more than 50 variables. It will give you all results anyways.

And the way that you can access that information, I am going to leave a couple of these paths open because we will revisit them later. But going back to census.gov/developers under the available API and right off the bat you can see American Community Survey one year. This is where you'll grab the example URL. In this case we want to make sure that we're grabbing an example URL that has group in it. But I want to work out of the data profiles again. So scrolling through the available options, once we get to data profiles in the ACS one year, you'll see the example call has question mark get equals group DP02.

So I'm going to copy that example and we'll work off of that. In this case I want to look at DP03 for all metro areas in the US. So pasting that example call we'll of course delete ampersand key equals your key goes here and press enter. It will give us all of the variable labels and then the values beneath it. It's not going to make a lot of sense on screen but we'll show you how to save this as a file. You just need to make a couple of edits first. We're going to change the year from 2018 to 2007 and press enter. Then we'll choose our table. We'll change it from DP02 to DP03 and press enter.

And then our geography portion, giving us data for the US by default, but we wanted all Metro areas.

So going to the example that we had visited earlier from the 2007 section of the ACS data profiles, on the left-hand side we see metropolitan statistical area as one of the options. The first example gives us data for all metropolitan statistical areas. And the second example gives data for an individual area.

I want all so I'm just going to copy the tail end portion that has the geography for the URL from the ampersand four equals metropolitan statistical area micropolitan statistical area wild card. And then once I copy that and go back to my URL here we'll replace the ampersand for US with this information and then press enter.

And then the last thing that you may want to do, one thing that we can add to any call - some of them are there by default in the example URL - is name. And that just gives you the label associated for all these metro areas in your results. So add a name just before group and add a comma and press enter. And here we have our results.

Once you get here, for any of the API calls you made you can right-click and save it as a CSV file and open it up into Excel. So I'm going to right-click and choose save as. In order to get it as a CSV, go ahead and start putting in the name. Here I'm going to put DP03 all metro 2007. And you do need to add dot CSV to the end of the file name and change the save as type to all files. And you'll be able to open it up and it will generally open up in Excel as a default. And here you get something that looks very similar to the downloads that you would get on data.census.gov. Each geography is its own row and all of the estimates are flat across that row without any special formatting or indentation.

With the Excel, you can do a couple of things to clean up the file here. And I'm just going to show you a couple of those. I have this all in step-by-step

slides. One thing that we noticed is there are some special characters for the brackets and the double quotation marks. If you use Excel's find and replace, you click find and select in the upper right and click replace, you can choose to find all instances of the open bracket. Leave replace with blank because we just want to delete them and select replace all. So now it found all 512 instances of that bracket and deleted it. Repeat this for the closed bracket and the double quotation mark.

Next, one thing you'll notice is the geography name because there is a comma in the label and it's separating based off of the commas, it split Galesburg, Illinois micro area into two columns instead of keeping it under the geography name. This has actually caused some of the data to be misaligned but you can fix this pretty quickly here.

So what we want to do is merge column A and B under name and shift everything back over one so that all of these values in column C appear under the label DP03_001E. So you right-click column C and insert. You can use a very quick formula in Excel. Just type equal into cell C2, click on Galesburg, type ampersand and click on Illinois micro area. So now we're telling Excel to merge these two cells together into one. We have that label Galesburg, Illinois micro area. In the lower right you'll see a green square. And once you double-click that, it applies it to all of the data down this row.

So now all we have to do is we want to delete the source data but before we do that we want to make sure that we locked down the values instead of relying on the formula to populate this information. So just a symbol copy-

paste would do this. Right-click column C and select copy, and right-click it again and choose the second paste option. When you hover over it it says value. So notice once we've done that and we're clicked into these cells, in the information above it does not show that formula anymore. It shows just the value, the label for that particular geography.

Once we've done that, we can go ahead and delete the source information and realign it. So selecting some of the information here from A and B and press control shift down arrow, we can right-click and delete that source information leaving the radio button to shift cells left. And now we're almost there. Notice we have our name under that one merged column but we have the extra cell C1 and we need to delete that as well just by right-clicking the cell, deleting it, and shifting the cells left. Now we have all the data realigned correctly.

So last step that you can do is pull in the variable labels. I'm going to walk through this really quickly. Know that there are step-by-step slides that go through this process that you'll be able to look on. But from the variables page where we found median income and owner for the 2007 vintage of the data profile, when you delete dot HTML from the end of that you can also save that just like we did for the other underlying data for all metro areas. And once you open this up, it gives you all of the variable labels and the description of what they represent.

So just like before, we can go ahead and remove the special character for both the brackets as well as those double quotation marks. And then we can do

control A and copy that information and put it in a second worksheet. So from here with this formula we can go ahead and grab those labels in. One way that you can do that is using vlookup. So equal vlookup and the open parentheses and then click on cell B1. This tells Excel that it wants to find a match for whatever content are in that cell. And then you put a comma in and you select all the information that we found earlier and press F4 one time and a comma. And then you just put in two comma zero and that tells Excel that you want to look for a match somewhere in this content and feed in the value from column B and that you want it to be an exact match.

Once you've done that, you should see if you've done it successfully a value populating here and then you can drag this across as needed to fill in. This will give you at least the labels associated with the estimates and know that the naming convention would carry over for the annotation for that estimate the margin of error, and the accompanying annotation for the margin error. Annotations just appear if there are special values.

Just a couple things to wrap up as we're getting ready to open up for questions. Again, there are step-by-step slides. And know that includes the formatting in Excel as well. Just to summarize benefits of using the API, it allows you to get the exact variables and geographies you need, you can quickly update your queries, and most variables are consistent from year-to-year, and it has more options to get data than on data.census.gov.

Some limitations - you can only request one geography summary level at a time. As an example, if you wanted data for California and all counties in

California, that would be two separate API calls that you would have to make. And not all collections are geographies are available. So there's not a way to get data in a single call for all Census tracts in the United States for other collection.

With those limitations in mind but also keeping in mind the benefits to kind of balance whether you may want to get your data from the API, data.census.gov, or the FTP site. Most queries and variables are consistent from year to year but you do need to check for changes. As an example from the ACS here, you can see in the first row for 2018 to 2016 you just need to update the year in the URL in order to get data for median income over time. But if you wanted data for the population under five years in 2018 the API call for that would be DP05_005E versus 004E in the 2016 data.

Variables can change for any number of reasons from the ACS. Any change in the underlying table can cause a change in the API variable name since they are code-based and table-based. Here you can see that variable change. If you made a call for DP05_004E in 2018 that of course gives you data from that table of the fourth row which would give you a value of 97 for the United States the sex ratio of males per female. But in 2016 it would give you a very drastically different value of over 19 million representing the population under 5 years. So do keep that in mind that you'll want to check for those changes.

With that, we'll start opening it up for questions. Operator, can you give the instructions?

Coordinator: 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 when prompted. Your name is required to introduce your question. If you'd like to cancel your question, you can dial star 2. It will take a few moments for the questions to come through. Please stand by.

Tyson Weister: Okay. Thank you so much. And as you're waiting for that, just wanted to have a few reminders here and additional information. We invite you to connect with us on social media using hashtag Census data. What we were able to show today is one piece of what's available in the full suite of educational materials for the API. If you visit the link in the upper left, you'll be able to access our API user guide, this recorded webinar, as well as step-by-step flyers with screenshots showing you more information on how to use the site.

And today is put on by our Census Academy Program. So not only do we have more resources for learning about the API, but anything you could think of related to Census data, tools, and how to access across all different topics, surveys, and programs. Our one-stop-shop for that information is on census.gov/academy. So we hope you check that out as well as a way to find more about our short videos and data gems, recorded webinars, and even full-length courses.

And that team as well can provide local training for you. If you visit census.askdata@census.gov with an email request or call the phone number listed here, they will be more than happy to provide a tailored, local training or give you local statistics. And we'll leave this up as we start answering some

of the questions. Do you have some queued up, Operator?

Coordinator: We do. Our first question today comes from (Robert). Your line is open.

(Robert): Hi. (Robert) here. Thank you for the great information. I have a question about - a couple questions actually. One, suppose you wanted to - I'm going back to your earlier example about homeownership - suppose you wanted to get homeownership rates for California for several years. Do you have to submit multiple APIs in order to do that or can you pull across your - so that's one question. And suppose you wanted to use ACS PUMS to get homeownership stratified by ethnicity or something along those lines. Can you also use APIs to do that?

Tyson Weister: So in terms of accessing the data over time, in general you will have to update the year. The only exception to that would be if you were pulling the data instead of the data profiles there is a product in the American Community Survey called the comparison profiles. In that instance, you could find statistics back to five years without having to change the year in the URL. So that table actually if you visit on data.census.gov and compare it to the data profiles will probably help you visually see this better with the data profile just be one year at a time and the comparison profile for that particular year would give you the most current for that particular vintage that you're clicked into as well as the previous five years.

In terms of the PUMS on the American Community Survey, right now we are transitioning in terms of getting that on the API for the first time. So if you go

to census.gov/mdat it's our microdata access on data.census.gov. And that of course just like everything else on the site is pulling from our API. We are still in the process currently labeled as a data site. So we are working on getting that in our public API.

And in terms of having information on how to access that just from the discovery tools, there isn't resources on that yet but once you work through the microdata access tool and create your table, you are able to click into downloads and that will give you the API call. So there's some functionality right now to start playing around with ACS PUMS on the API But we're still working on getting more documentation available for later.

(Robert):   Thank you.

Coordinator: Our next question today comes from (Amy). Go ahead. Your line is open.

(Amy):   Hi. I was calling - wanting to figure out. I do a lot of research with American Indian tribes. And trying to figure out how to best access older data from the decennial censuses 1990 and 2000. I've had success pulling the information from the ACS five-year averages. But wondering will those geographies be available sometime soon?

Tyson Weister: Yes. Let's double-check here. What I know about decennial census data offhand is that there are currently a limited number of geographic levels that you can access the historical information for on the APIs. We do have 1990 and 2000 on the API but it's only for a limited set of geographies.

With that, we are working to add the additional geographies. That's part of our process in getting the data on data.census.gov for the 2000 census. So I'm going to start with the 1990 here. Again, I just recommend going into the examples and we want to get to the decennial dataset. I'm just going to see if I can pull that up quickly. Also another good keyword is SF1, summary file one. Let me just double-check here to see what's available. Okay.

So I'm not showing that as currently on the API in summary file one. So most of the AIAN data you would want to use the FTP site for, as we work to migrate at least Census 2000. But you would still be needing to go to the FTP site for the 1990 data.

(Amy): Thank you.

Tyson Weister: Operator, do we have questions queued up?

Coordinator: We do. Our next question today comes from (Vince). Go ahead. Your line is open.

(Vince): Hello. The answer to this question may have been implied by your answer about PUMS from the first caller. But I really loved the old Data Ferret and the flexibility to be able to create cross tabs that might not have been created by the Census in their existing tables. Is that sort of functionality available through the API or is it reliant on the sort of row and column design of the tables that the Census has already created?

Tyson Weister: Sure. So the API functionality that I showed in this webinar today is all based on our pre-made tables. The microdata is where you would want to go now. So the data ferret has been taken offline but we have new tools to access that information as part of data.census.gov. When you go to data.census.gov scroll to the bottom and click on microdata access. Or just go to data.census.gov/mdat. This is where you can navigate to the different data sets and select variables to create custom tables that you can't find in the set of pre-made tables.

And once you go through this process, what I was referring to earlier, generally folks would often just kind of view their results on your screen but once you go to download you do have some options to get the data from this API call whether you wanted the underlying raw data from the API or the API query that would tabulate the results as you were seeing on screen just a moment ago.

And there's more information on using this particular tool. If you go to what is data.census.gov, we are the enterprise dissemination for the Census Bureau. So that includes the pre-made tables on data.census.gov, custom tables you would want to create using the microdata access, as well as the API. And under the guidance for data users section, the link in the webinar PowerPoint I provided was how-to materials for using the Census API. But we also have how-to materials for using microdata access.

(Vince): Excellent. Thank you so much.

Tyson Weister: You're welcome. Operator?

Coordinator: Thank you. Our next question today comes from (Jill). Go ahead. Your line is open.

Tyson Weister: I'm having a hard time understanding hearing the question come through.

Kim Davis: (Danielle) can we please move on to the next question? We're having trouble hearing this particular caller. But you are welcome to queue back by pressing star 1 on your phone. Thank you.

Coordinator: Yes. Our next question comes today from (Anna). Go ahead. Your line is open.

(Anna):   Hi. My name is (Anna Unda) and I would like to know if we can retrieve different age groups from those who are already set-up by the tables. For example, if I want a different age group 18 to 25 and define my own age groups, can I do that?

Tyson Weister: You would want to do that either through the PUMs on the microdata that I had referenced earlier where you create your custom tables. But we also do have single year of age at the Census Bureau through the Population Estimates Program. So you may want to check that out as well.

(Anna):   Thank you.

Tyson Weister: You're welcome.

Coordinator: Our next question today comes from (Zoey). Go ahead. Your line is open.

(Zoey):   Hi. I was curious if we can select data using like a custom geography or do we have to use the pre-staged ones that have already been set up like states and counties?

Tyson Weister: The pre-made geography are the only options that are available right now. But it does cover down to the block group level from the American Community Survey or the block. So what some folks do when they want to get data for a custom geography is get data for one of the small level geographies that are available and then match up as fast as they can and aggregate in order to create data for their custom geography - even though it may not be exactly aligned at the boundary but it may give some sort of estimate. That would be the best recommendation that I would have for you.

(Zoey):   Okay. Thanks.

Coordinator: Our next question today comes from (Bob). Go ahead. Your line is open.

(Bob):   Yes, hi. Thank you for this great presentation. Two questions, if I could. The first one I think is fairly simple and it's if I want data for multiple block groups, how would I format the query or do I have to have individual queries for each block group? And the second question is, you know, it's great to get

the numbers but it's more important sometimes to understand the numbers. So, what access do we have to data describing the metadata? I'll just give you an example. I'm interested in fuel use in housing. When I look at the data, I get for instance in New York, a lot of houses are being heated with coal. I think the number are high.

I'd like to know your source of information is so I can understand when I see estimates for a number of houses what that means. Where would I look for that? So, there are two questions. How do I get multiple bar groups and otherwise, where do I get metadata on the data that's being returned to me?

Tyson Weister: Those are great questions. So, actually I have one of the examples pulled up here. This particular data set doesn't go to the block group level that I'm clicked into, but the concepts would be the same. So, I'm going to run the query here first to get data, narrow down to a specific state. And then, to get data for one county, we would just do 181 as one example. And then, if you just type a comma, so I had the code 48 and then under County for the first County, I had to type the County Code 181.

And then I just put a comma in and then put in 183. You should be able to do something similar at the block group level but you do have to work within the hierarchy as well. So oftentimes it's block groups within a census tract and things of that nature. I'd have to look at the structure more but that may be a good question. If you want to put it in the Chat so I can follow-up with, with more specific information, for that particular geography. But in general, you should be able to type in the comment and add additional codes. In terms of

your...

(Bob):   Is there a limit to how many things I can put in with the comma there? Block groups are small but I typically run like 100, 200 at a time. But.

Tyson Weister: I'm not certain what the limit is offhand. We could also follow-up with you on that.

(Bob):   Okay.

Tyson Weister: In terms of the documentations here on the different parts. I'm going to go back from the navigation here. The American Community survey has subject definitions and actually I'm just going to navigate to the American Community Survey website. So, you want to make sure to look at the documentation for the different surveys and programs. The American Community Survey is really great at documenting on their own Website. If you go to the technical documentation side, on the left navigation, and then the Code Lists, Definitions, and Accuracy, and under the subject definitions.

So, this is a 162-page document on all of the different topics in the American Community Survey in definitions. And you'll notice there's a section for health teams so you can click there and then read more information. At the beginning it will also tell you, in the American Community Survey form, what the question number was that the data came from. So, you can even pull up the sample form and view how the question was asked, what the response categories were. And then you'll of course get detailed information below it as

well.

Kim Davis: Okay. So, we have hit and actually gone past the hour. This is your host Kim Davis and we won't, unfortunately, be able to complete any further questions from the Chat or on the phone line at this time. But we are happy to follow-up with any of you that entered anything in the Chat and was not able to get your question answered. You can follow-up with us on the, with the information on this screen. Or, by emailing census.askdata@census.gov. Or, by calling 844 ASK DATA and we'll be happy to help you and follow-up.

We want to thank everyone for participating today and thank you, Tyson, for presenting for us today. This is a wonderful presentation and very helpful for many of us. Thank you all and have a nice day.

Coordinator: That concludes today's conference. Thank you for participating. You may disconnect at this time.

END