American Community Survey (ACS)

Using Census API with the American Community Survey Webinar Transcript

August 15, 2016

Javier Gomez

American Community Survey Office

This presentation was presented as a webinar to the General Public on August 15th, 2016 by Javier Gomez of the U.S. Census Bureau's American Community Survey Office. A link to the recording is available at https://census.webex.com/census/ldr.php?RCID=2c135ca736678df8c4ad2c58e770c988 The transcript of the webinar follows. Slide references and links have been added to the spoken text as appropriate.

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 that time to ask a question, please press Star 1 on your touch-tone phone and record your name at the prompt. This call is being recorded. If you have any objections you may disconnect at this time.

I would now like to turn the call over to Mr. Javier Gomez. Sir you may begin.

Slide 1 – Title Slide

Javier Gomez: Thank you. Good afternoon, my name is Javier Gomez. I work for the American Community Survey Office, Data User Education Branch. Today I am going to be presenting how to use the Census Application Programming Interface or API, with the American Community Survey, or ACS.

Slide 2 – Outline

The outline for this presentation will be as follows. I'm going to start by talking about the ACS basics and content, then Census API basis and data sets available. We will follow that by talking about what API URL looks like and how to use and construct an API URL. And finally information about how to stay in communication with us and questions.

Slide 3 – ACS Basics

The American Community Survey is an ongoing National Survey that samples approximately 3.5 million addresses annually which comes out to about 290,000 addresses per month. The data is collected continuously throughout the year to produce detailed population and housing estimates each year.

So you can think of the ACS as a complement for the decennial census. The decennial, as it implies in its name, is conducted every ten years with the next one coming in 2020, and that one is sent to the whole population in the country. On the other hand, the ACS is a survey that is sent to only a sample of the population. It was fully implemented in 2005 as a replacement for the long form every ten years.

So if you remember a few years back the decennial census used to contain a short form which was sent to the whole population and a long form which was sent to a sample of the population but this was conducted every ten years. So in order to have more timely data the ACS substituted the long form, as I mentioned in the year 2005. It is designed to produce critical information in the small areas in a small population groups previously collected on the decennial long form. It covers over 35 different topics and supports over 300 known federal government uses and countless other nonfederal government uses.

We released two different sets of estimates each fall, 1-year estimates which are released in September, and 5-year estimates which are released in December. I am going to elaborate a little more on that later on in the presentation.

Slide 4 – ACS Content

Some of the information that is available through the American Community Survey contains social data or variables such as ancestry, citizenship status, education, fertility, immigration. We also obviously have the demographic data like age, Hispanic origin, race, relationship. Economic data like class of

worker, commuting to work, health insurance, income. And finally, we have housing data with variables like computer and Internet use, home value, utilities, and vehicles.

As mentioned before, the ACS covers over 35 different topics from where we produce over thousands of tables and more than 11 billion estimates.

Slide 5 – Census Geographic Concepts

On the screen, you can see some of the geographies for which ACS data is produced and the relationship between them. Lower geographic areas fit neatly within the larger areas directly connected with lines. For example, the school congressional and state legislative districts fit neatly within states and they do not cross state boundaries. However, they might cross boundaries of other counties in metropolitan areas. Statistical geographic areas are defined primarily for data tabulation and presentation purposes.

As you can see on your screen, the smallest geographic areas for which we publish data on the ACS is block groups. Block groups are statistical divisions of Census tracts and are defined to contain a minimum of 600 persons or up to 240 units. In total, we have over 930,000 geographic areas for more than 35,000 communities throughout the country.

Slide 6 – Availability of ACS Data Products

Now the availability of ACS data products. They are released one year after the data is collected, and the first year of data collection with full sample was in the year 2005. Therefore, the ACS data that was collected in 2005 was released in 2006 as one-year estimates.

ACS five year estimates pulling data over, collected over 60 months between 2005 and 2009 were released for the first time in 2010, one year after the fifth full year of ACS data collection. ACS five-year estimates were then released each year using the five most recent years of data. ACS data is available for geographic areas with a population of 65,000 or more in the form of one year and five-year estimates. ACS data is only available for geographic areas were population is less than 65,000 in the form of five-year estimates. The Census

Bureau discontinued the three-year estimate notice to maximize efforts on the one and five-year estimates. Previous three-year data products are still available through the Census Bureau's website.

Slide 7 – Outline

So now that we covered the basics, or an overview of the American Community Survey, let's move into the census API basics and data sets available.

Slide 8 – API Basics

First of all API which it stands for application program and interface is a tool for collecting the variables and data that you need in raw format from the Census Bureau's ever-increasing pool of data sets. It is set to represent data in the standardized way, and by standardizing the API credit language across multiple data says the learning curve for developers is reduced.

Researchers and developers on both inside and outside the census use that API because of the simple format. It provides greater ease for inputting, presenting, and manipulating data in whatever format you choose. With an API you can select any variables in multiple data sets and place them in raw format, for the format of your choice.

Slide 9 – API Basics

Some of the benefits of using the census API include that you can query data directly from the census servers. There is no complex interface. You can download only the variables and the geographies that you need, which I think is something important because there is no need to download a huge data set if you only need a few variables or geographies of that data set.

Another benefit is that you can connect to a statistical analysis software such as R. There is immediate access to the most current data. There is no need to host data in another server, once again something important because as I mentioned before you do not need to download the huge data sets. You can

just connect to the server, download, and retrieve the variable and geography that you need. It is widely used to build mobile and web applications and drives interactive data visualizations.

Slide 10 – Examples of Apps that Use the Census Data API

One example of applications that use the Census data API is the Cornell University Poverty Application, which retrieves data simply from a different county as is displayed on your screen.

Slide 11 – Examples of Apps that Use the Census Data API

Another application that uses the Census API is the population clock on the Census webpage,

Slide 12 – Census Web Applications

or quickfacts, and my Congressional District.

Slide 13 – Datasets Available

This presentation is focused on how to use the API with ACS data but the Census hosts a number of different data sets. The most popular aggregate data sets to quite a few variables using the API are, once again obviously the ACS estimates with one, three or five years products, supplemental tables, summary files, data profiles, and also we're including in the upcoming release comparison profiles and subject tables.

You can also find data sets for Economic indicators time series, Decennial Census, Economic Census, CBP and Non-employer statistics, population estimates and business economic statistics. There is a complete lists of all of the data sets available at the link api.census.gov/data.

Slide 14 – Outline

Now we are going to cover one API URL looks like.

Slide 15 – What an API URL Looks Like

An API URL pretty much looks like a regular webpage URL. In this particular, example which I'm going to cover step-by-step along the presentation, but in this particular example we are requesting a Web browser to go to the Census servers and to take 2014 ACS 1-year estimates data from table B01001 for all states. So it is pretty much like going to fact finder, find table B01001, and then get the total population estimates for every single state.

Slide 16 – Build the URL

How to build the URL. The first step is to identify the data set that you want to use. As I mentioned before, there is a number of different data sets the Census has data for, so that would be your first step to identify which data set that you want to use.

Second step - identify the items that you need, which variables you want, and for which geographies you want them. Third, build your query, and in order to build your query you will the functions "?get", for the variables and names separated by commas, and the "&for", for the geographies to specify for which geographies you want this data, and then put it all together.

Slide 17 – API Data Sets

Before we look into that step-by-step, here is is an example of the different data sets that we have available at the census.gov webpage. As I was mentioning before we have the American Community Survey data set with 1-year data, 3-year data, 5-years data, and supplemental data now.

Slide 18 – API Data Sets

Here is an example of how if you click on one of the data sets how it would look. This particular example is ACS data 1- year, so you can see 2014 data set, 2013 data set, and you can find examples of the variables available in that data set, and technical documentation as well as the supported geography for that data set in particular.

Slide 19 – ACS Variables

Here is an example of the different variables that you can find in a data set.

Slide 20 – Outline

How to use and construct an API URL.

Slide 21 – How To Do a Query

The first step is to start your query with the host name http://api.census.gov/data. All your API URLs are going to start with this host name.

Slide 22 – How To Do a Query

The second step is to add the data years to our URL from which they are - what is the year of the data set that you are interested in? For this particular example, we are choosing 2014 data set.

Slide 23 – How To Do a Query

The next step, the "acs1" stands for ACS data 1-year estimates. This is a data set, the name of data set that we want to choose. As I mentioned before if you go into the Census webpage you can find a whole list of data sets available.

Slide 24 – How To Do a Query

Now at this point all we are telling the browser is to connect with the census server and get data from the 2014 data set ACS 1-year estimates, but we haven't told our browser what data we want, what variables, which variables and for which geographies. The query starts after the question mark.

Slide 25 – How To Do a Query

Then we use the function "get" to add all the variables that we want to retrieve data for. All the variables that you list have to be separated by a comma. In this specific example, we are requesting a name and then data from table B01001. The underscore 001E, that stands for total population. Once again, if you look at the table with all of the variables you will find the specifics of this variable.

Slide 26 – How To Do a Query

Finally, we use the "&for" function to specify for which geographies we want the data. Here we are using the wild card colon asterisk, which means to get the data for all the states or all the geography available.

Going back here, if you want to specify a state you just input the number of the state, and if you want several states you can just separate them by a comma.

Slide 27 – Output

The results from the census API data queries will be return in JSON, and that data can easily be plugged into the format and file or graphic of your choice. JSON or JavaScript Object Notation is an easy format for exchanging data between platforms using human readable language.

The Census uses is a non-standard version of JSON that is streamlined. Data is represented in name, value pairs, square brackets, full arrays and values are separated by comma.

Live Demo

At this point I am going to switch a little bit to a browser and show you how this would work. If we take this URL and put it into a browser, we retrieve the data in JSON format. So what we were saying here is to retrieve data from the 2014 data set, ACS 1-year estimates, and variables name and total population from table B010014 for all the states.

And so we can play around with this URL. Let's say that instead of all the states, we want for the states of District of Columbia, Maryland and Virginia. So here we're just requesting the data for those three states. Then if instead of 1-year estimates we want 5-year estimates, we just change from acs1 to acs5.

Slide 28 – Register for a Key

I am going to go back to the PowerPoint presentation. You can include up to 50 variables with a single API query and up to 500 queries per IP address

without registering for a Census key. In general, this is used for testing purposes. However, if you are part of a group of people at an IP address making queries, such as a college class, you can hit that 50 variable limit quickly and you should request a key. When you are ready to pull this data to a web site register for a key and that key will be part of your data request URL string.

Slide 29– Resources

At this point, I would like to share additional resources. There is a data set discovery tool at api.census.gov/data. Census developers webpage at census.gov/developers, and how to convert JSON to CSV data at the link that you can see on your screen.

Slide 30 – Outline

Slide 31 – Source Us!

How to stay in communication with us. I would like to invite everyone who uses the data to source us. Just include us in your publications every time that you use our data.

Slide 32 – Continue the Conversation #ACSdata

Continue the conversation by using the hashtag ACS data throughout the different social platforms such as Facebook, twitter and YouTube.

Slide 33 – American Community Survey Data Users Group

Also, we want to remind you that there is a users group specifically for users of ACS data. The data users group already has over 1400 participants and is free to join, and there is an API group at the link that you can see on your screen.

Slide 34 – Need Local Stats?

As a reminder, we have data dissemination specialists throughout the country who can provide data workshops locally. If you are interested in a workshop, please contact clmso.dbbquestions@census.gov.

Slide 35 – Questions

That will conclude my presentation. At this point we can take some questions. Thank you.

Coordinator: Okay thank you. And if you'd like to ask a question please press Star 1 on your touch-tone phone, unmute your phone and state your name clearly at the prompt as your name will be required to introduce your question. To withdraw your question please press Star 2. One moment please for our incoming question.

Thank you. Our first question is from (name removed). Your line is open sir.

(Question #1): Thank you Javier.

Javier Gomez: Hi. How are you?

(Question #1): Good, thank you sir. My best guess is the answer to this is no but let me go ahead and ask. Via this API tool, is it possible to query like custom table selects where you are creating your own with that would be a custom cross tab or can you strictly only query results for tables as they are now?

Javier Gomez: One moment please.

(Question #1): Okay.

(Bill Hazard): Yes hi. This is Bill Hazard. My group is responsible for the API. These are all aggregated estimates that are coming back to the API so you can't create any special cross tabulations. You're stuck with the estimates that you, you know, that have been created already. You can however - you're not stuck with just the variables from a single table in a query so you could pull specific cells or what you would consider rows from different tables in one query.

(Question #1): Got it. So where data ferret will allow you to in fact create custom cost tabs the API will not but it sounds like you can query more than one table at one time to get one result?

(Bill Hazard): Right yes. So yes I mean, data ferret is using the - if you creating a custom cross tabs you are using the micro data, the public use micro data.

(Question #1): Right okay.

(Bill Hazard): Yes. So in this case, there is - we will have at least the plan is to have make micro data available through the API at some point where it's actually creating the tabulations via the API so at that point I guess you would be somewhat creating custom cross tab.

(Question #1): All right.

(Bill Hazard): …the big idea like how you define a cross tab via an API but…

(Question #1): The micro data is not available to interface this yet though?

(Bill Hazard): Not currently.

(Question #1): Got it. Javier are you going to make the PowerPoint presentation available?

Javier Gomez: Yes. It will be available soon.

(Question #1): Okay. Thank you very much.

Javier Gomez: Okay.

Coordinator: Thank you. Our next question is from (name removed). Ma'am your line is open.

(Question #2): Hi. Yes I'm wondering if it's possible to specify the fields that are returned because I notice on your query it didn't give us the margin of error field and I'm also wondering specifically about a Geo ID field so you can map this stuff?

(Bill Hazard): So yeah, any variable that is available can be queried. So you would just, you know, basically find the variable name in the list of variables. So, in a case like this, it would basically be the same thing but replacing the E with an M so you get the margin of error for that, and I believe Geo ID is also a variable that is on the data set so you could request that also.

(Question #2): Okay great thanks.

Coordinator: Thank you, and our next question comes from (name removed). And again, if you'd like to ask a question please press Star 1. (name removed), your line is open please.

(Question #3): Hi. This is (name removed). I didn't get a clear answer to where can I get a copy of the slides shown in this presentation? It went by pretty fast and I was hoping I could get them and print them out and look at them with care.

Javier Gomez: Yes, we are going to be sharing a link where you can get the - a copy of the presentation.

(Question #3): Where will I look for that? Where will I get that? How will I get it?

Javier Gomez: On the email, on your email.

(Question #3): Oh okay. Thank you.

Javier Gomez: You are welcome.

Coordinator: And thank you. We do have one other person who has joined for a question but they didn't announce their name so if you just ask for a question your line is open.

(Question #4): Hello?

Coordinator: Go ahead.

(Question #4): Hi. Is GeoJSON access through the API or Census Blocks?

(Bill Hazard): There is a separate geographic API. So on the census.gov/developers page you'll see there are - there's all these data sets that are what we call the data API. Then there's two - there are two geographic APIs. I'm really not that familiar with them but they I think are probably what you're looking for.

(Question #4): So I go to the API page and I look for the geographic APIs and…

(Bill Hazard): Yes.

(Question #4): …that'll - okay thank you.

(Bill Hazard): Yes.

Coordinator: Thank you, and currently I'm showing no further questions in the queue. Again if you like to ask a question please press Star 1. And there are no further questions at this time.

Javier Gomez: At this time if you have a chance I would like to just ask you to fill the Training Evaluation Form it doesn't take more than two or three minutes.

Coordinator: And excuse me we do have two more questions jumping in. One moment.

Javier Gomez: Sure.

Coordinator: Our next question is from (name removed). Your line is open sir.

(Question #5): Javier, you had a link in the presentation of convert JSON to CSD data. Just to get a heads up on that is that all that complicated or fairly simple? How tricky is that?

(Bill Hazard): I don't know. I don't - I've never used it myself actually. I don't think it's that tricky and it's not even that hard to bring in just the regular JSON output into Excel just because it's delimited.

(Question #5): Got it. So are folks using this whole API method in a more simpler fashion than creating applications just to get the data in a different way…

(Bill Hazard): Well I mean…

(Question #5): …made more efficient?

(Bill Hazard): You know, it is intended for developers that are creating applications right so they can yes so they can pull that data dynamically. We have found though that sometimes is a very quick and easy way to basically create a download or an extract of a piece of a data set, you know, and then pull it into something like an Excel or even SAS

(Question #5): Right, kind of a streamlined version of either Data Ferret or AFF itself.

(Bill Hazard): Yes where if you just want to pull down the raw numbers and be able to manipulated yourself.

(Question #5): Got it, thank you.

Coordinator: Thank you and our next question is from (name removed). Sir your line is open.

(Question #6): Hi. Yes this is (name removed).

Javier Gomez: Hi.

(Question #6): I had a question well, maybe there's not a lot of questions right now because, you know, well maybe for a lot of people this is new to them. But if we once, you know, we start working with the API, I could imagine that I'll have a lot of questions and where should I go to get those answered?

(Bill Hazard): There should be links on the Developers page for help. It is actually the Data Ferret helpdesk that actually sit in the same branch, they created the API and maintain data ferret. And I'm just trying to think what link is on the web site but there should be resources there for getting assistance.

(Question #6): Okay so I mean is it, and I haven't even gone to it but, is it sort of like FAQs or is it information that you can read the maybe with examples…

(Bill Hazard): There are definitely examples all throughout. I think that is one of the easiest ways to start using it. So Javier showed like the - some screenshots of the like the data sets, data html page, and when you dig down into, you know, click on into the details of the data set you see a link for examples, and then, for every geographic level available for that data set you have two to three examples. They just basically put one or two variables in to get but you see the syntax and it's a real easy way to start using it. There are also some more somewhat maybe more complex examples listed on the developer's webpage that - so the two things in combination can get you a lot of information.

(Question #6): Okay thanks.

Coordinator: Thank you. We do have one more question. One moment please while I get their name. And our next question is from (name removed). Sir your line is open.

(Question #7): So actually I guess maybe the previous answer sort of answered this but it would be great to have some more examples. So I guess the developers webpage has more examples of how this has been used?

(Bill Hazard): How it has been used or actual examples of…

(Question #7): Well actually examples of, you know, yes actual examples. It would be great to sort of see what other people have done with this just too sort of, you know, give us a sense of what…

(Bill Hazard): Right.

(Question #7): … What are some of the possibilities?

(Bill Hazard): You know that is a good idea. I am not sure that we really showcase much. There were - so we had a - there was an API forum up but it might still be there although there were issues with it still being active. Like I am not sure you can continue to join but that was a place where we were trying to get people that were using it to put in there, you know, what they've been doing in different things.

So if that link is still alive you might be able to find some information there and some examples of what people are doing, but we are working on trying to improve the developers web site so that we do provide more information, and obviously that would be a good example or a good idea to add some examples of how it's being used.

(Question #7): Yes. I guess the other question I had of the little unclear. So is the data are we - are you when you use this functionality is the data actually being downloaded or is it, you know, working with that or is it actually taking it live off of the Web - off the Census servers and you are working with that with the live data?

(Bill Hazard): It is pulling it directly. Each time you put a URL in it pulls the data directly from Census servers. So in the case of if you wrote an application that is basically a dynamic application every time you make any kind of a, you know, change to that to that application it's going making the call to the data at Census and pulling that back to present it to the interface.

(Question #7): Okay. So you can - okay so if we were to develop something that was working with the 2014 ACS data and once the 2015 is available we can just change the call to 2015 and it would just grab that same variable and…

(Bill Hazard): Yes, exactly.

(Question #7): …without having to download all the time?

(Bill Hazard): Assuming that variable still exists and didn't change, you know?

(Question #7): Okay.

(Bill Hazard): Yes that's yes exactly right.

(Question #7): So how consistent are the variable names? I mean haven't run any issues in the past but I always worry about that.

(Bill Hazard): Right, right. Well we are certainly trying to promote consistency now. I mean, I think it's becoming over the past, you know, few years or how many years it's become, we've sort of stress that and, you made - tried to impress upon data providers that it's an important feature that data users want to have, consistency over time, so the more we can keep things the same the more we - we're doing that.

(Question #7): Are there any sort of warnings like they do within the ACS in terms of the PUMS data that, you know, there were some key, you know, changes made to these variables. Is there anything like that for the variable level data from the

ACS - or the Factfinder or would you get a warning that that variable isn't - doesn't exist anymore or if that - or this variable has changed or…

(Bill Hazard): Well that is one so, I me one of the things we can work on is improving the information that is presented about each variable. We have basically all that information is being stored and maintained in the metadata repository. So we certainly have it. I guess in the case like several people have mentioned data Ferret that controls use of variables that are comparable over time.

So in the same metadata repository the same idea it's being used here. It's just right now we probably aren't presenting that through the API documentation but it's there. So it's just a matter of, you know, as we write additional documentation, you know, presentation for the API then that information would be there available. We wouldn't be able to control that, you know. The use of it is not really an interface but we could present the information on what is comparable to where things changed.

(Question #7): All right great. Thank you very much. Appreciate the presentation.

Coordinator: Thank you, and currently there are no further questions at this time.

Javier Gomez: All right.

Deborah Rivera-Nieves: Okay this is Deb Rivera. I am here to facilitate the training and help out Javier during his presentation. Just as a final note, I want to remind everybody that I will be providing the WebEx recording link. That will be sent to the email address that you used to register for this presentation today. So once again, we want to thank you for joining us today. If you have a minute please fill out the evaluation form and once again thank you for joining us.

Coordinator: Thank you for participating in today's conference. That does conclude this call. Please disconnect your lines and speakers standby for post conference.

END