

Getting Started with the Census Data API in R Using *Tidycensus*

Catherine Tulley (she/her)

Transportation Planning Data Analyst, SPC

March 9th, 2022 @ [PA GeoDev](#) | Virtual Breakout Session 1, Track 3 | 10:15 AM – 11:00 AM



Visit Catherine's GitHub repo for the exercise files: https://github.com/cjtulley/PA_GeoDev_2022

Quick poll

Go to www.menti.com and use the code 7158 3051



What brings you here today?

Mentimeter

- 1st | Old hat with the Summary File?
- 2nd | Used American FactFinder or data.census.gov?
- 3rd | Used ESRI's Census layers online?
- 4th | Just looking to learn more!

Today's Agenda

In this session, we will learn how to **query, download** and **format data** from the US Census Bureau's API (Application Programming Interface) in R, using Dr. Kyle Walker's 'tidycensus' library (plus a few other helper packages).

We will save the results to a CSV, to a formatted Excel file and to a shapefile – all using R!

Pt. 1: Intro to the Census API

The 5 W's

API Basics

Why Tidycensus?

Pt. 2: Dive into R & Tidycensus

stuff

things

stuff

Pt. 3: Resources + Q&A

Resources

Q&A

Thank You & Contact Info

Pt. 1: Intro to the Census API

Covering the 5 W's: What, Where, Who, When, & Why
(We'll get to the 'how' in a bit...)



Find out *what* you need to know...

Always a GOOD IDEA to start with the USCB's Developers page:

<https://www.census.gov/data/developers.html>

1. API Gallery – “Is there an API for my dataset?”
<https://www.census.gov/data/developers/datasets.html>

2. Refer to the API User Guide:
<https://www.census.gov/data/developers/guidance.html>

Tip: Scroll to the bottom for Video Tutorials & Webinars
(great guided intros to using the APIs)

3. Sign up for your API key ->



Need an API key?

Request it here:

https://api.census.gov/data/key_signup.html

...to get to *what* you need: the data!

Where can I find help? Who can help me?

A few suggestions:

- Check out good ol' Stack Exchange! <https://stackoverflow.com/>
 - Tags: R, Census
- Join the **US Census Bureau's Slack Channel** (*see Resources slides*)
- Contact the **US Census Bureau's CEDSCI Division** (Center for Enterprise Dissemination Services and Consumer Innovation) for API support by email:
 - cedsci.feedback@census.gov (CB staff generally respond quickly)
- **ACS Data Users Group Forums:**
 - <https://acsdatacommunity.prb.org/>



What are we doing & why are we doing it?!

1. WHEN you need more efficient means to access Census data!

- If you've ever spent WAY too long trying to download data from AFF (old) or data.census.gov (new)...and then reformatting it for Excel or GIS. 🤔
- If ESRI's Living Atlas layers aren't what you need
- If you need to create custom summary tables AND maps to go with them...

2. Send a query to the API

3. Get the data returned to you in a format that you can easily slice, dice and manipulate in R...

4. ...and output into any format that you might need, all from the same interface! 🥰

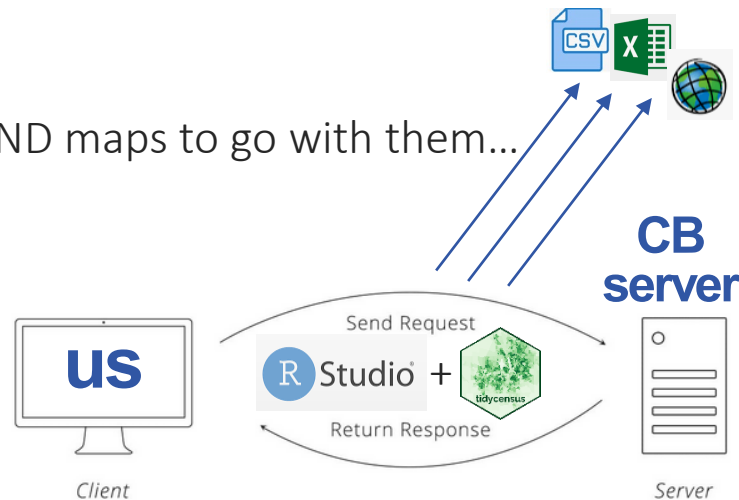


Figure 1. The Request-Response Cycle.

Graphics by Zapier, R Studio, Walker Data, Microsoft, ESRI, Twitter & Flaticon

API Basics – How to use the ACS API

1. We need to understand how it works before we dive into R

- Visit: <https://www.census.gov/data/developers/data-sets.html>
- Example: **5-year ACS data** -> Click the 5-year link to get here:
- <https://www.census.gov/data/developers/data-sets/acs-5year.html>

EXPAND ALL | COLLAPSE ALL

American Community Survey (ACS)	
November 30, 2021	American Community Survey 1-Year Data (2005-2020) Areas with populations of 65,000+. Covers a broad range of topics about social, economic, demographic, and housing characteristics of the U.S. population.
October 15, 2020	American Community Survey 1-Year Supplemental Data (2014 - 2019) High-level detailed tables tabulated on the 1-year microdata for geographies with populations of 20,000 or more.
September 15, 2016	American Community Survey 3-Year Data (2007-2013) Areas with populations of 20,000+. Covers a broad range of topics about social, economic, demographic, and housing characteristics of the U.S. population.
December 10, 2020	American Community Survey 5-Year Data (2009-2019) Data available down to the block-group level. Covers a range of topics about social, economic, demographic, and housing characteristics of the U.S. population.

2. Decide which vintage you want to use:

Comparing American Community Survey Data

ACS has non-overlapping datasets that allow comparisons of current ACS data to past ACS data. The 2015-2019 ACS 5-Year estimates can be compared with 2010-2014 ACS 5-Year estimates. For information on comparability of the 2015-2019 ACS 5-Year estimates to the 2010-2014 estimates by topic, please visit the [Comparing 2019 American Community Survey Data](#) page.

2019	2018	2017	2016	2015	2014	2013	2012	MORE ▾
------	------	------	------	------	------	------	------	--------

2019

Detailed Tables

- **Example Call:** `api.census.gov/data/2019/acs/acs5?get=NAME,group(B01001)&for=us:1&key=YOUR_KEY_GOES_HERE`
- 2019 ACS Detailed Tables Variables [[html](#) | [xml](#) | [json](#)]
- [ACS Technical Documentation](#)
- [Examples](#)
- [Supported Geography](#)

Let's take a look

Detailed Tables: Variables

1. You will need to understand these variables, because we'll be sending these to the API through Tidycensus.
2. Variables list: <https://api.census.gov/data/2019/acs/acs5/variables.html>
3. Example variables table – let's say we are looking at broadband availability:

VARIABLE	VARIABLENAME	TABLE NAME
B28002_001E	Estimate!!Total:	PRESENCE AND TYPES OF INTERNET SUBSCRIPTIONS IN HOUSEHOLD
B28002_002E	Estimate!!Total!!With an Internet subscription	PRESENCE AND TYPES OF INTERNET SUBSCRIPTIONS IN HOUSEHOLD
B28002_003E	Estimate!!Total!!With an Internet subscription!!Dial-up with no other type of Internet subscription	PRESENCE AND TYPES OF INTERNET SUBSCRIPTIONS IN HOUSEHOLD
B28002_004E	Estimate!!Total!!With an Internet subscription!!Broadband of any type	PRESENCE AND TYPES OF INTERNET SUBSCRIPTIONS IN HOUSEHOLD
B28002_005E	Estimate!!Total!!With an Internet subscription!!Cellular data plan	PRESENCE AND TYPES OF INTERNET SUBSCRIPTIONS IN HOUSEHOLD

The "E" stands for Estimate
An "M" would stand for MOE

The first part before the "_" is the table ID.

TIP: Always review the latest ACS Technical Documentation page to check for updates to the data between vintages (data release years). [link](#)

What does this look like from the Summary File? (think “Table Shells”)

TABLE ID:		B28002																		
SEQ NUM:		013500000000																		
CONTENTS:		2015-2019: 5-YR ESTIMATES FOR PENNSYLVANIA (CENSUS TRACTS AND BLOCK GROUPS)																		
TITLE:		PRESENCE AND TYPES OF INTERNET SUBSCRIPTIONS IN HOUSEHOLD																		
UNIVERSE:		Households																		
FILEID	FILETYPE	STUSAB	CHARITER	SEQUENCE	LOGRECNO	Geography ID	Geography Name	B28002_001	B28002_002	B28002_003	B28002_004	B28002_005	B28002_006	B28002_007	B28002_008	B28002_009	B28002_010	B28002_011	B28002_012	B28002_013
								PRESENCE AND TYPES OF INTERNET SUBSCRIPTIONS IN HOUSEHOLD%Total	PRESENCE AND TYPES OF INTERNET SUBSCRIPTIONS IN HOUSEHOLD%Total	PRESENCE AND TYPES OF INTERNET SUBSCRIPTIONS IN HOUSEHOLD%Total	PRESENCE AND TYPES OF INTERNET SUBSCRIPTIONS IN HOUSEHOLD%Total	PRESENCE AND TYPES OF INTERNET SUBSCRIPTIONS IN HOUSEHOLD%Total	PRESENCE AND TYPES OF INTERNET SUBSCRIPTIONS IN HOUSEHOLD%Total	PRESENCE AND TYPES OF INTERNET SUBSCRIPTIONS IN HOUSEHOLD%Total	PRESENCE AND TYPES OF INTERNET SUBSCRIPTIONS IN HOUSEHOLD%Total	PRESENCE AND TYPES OF INTERNET SUBSCRIPTIONS IN HOUSEHOLD%Total	PRESENCE AND TYPES OF INTERNET SUBSCRIPTIONS IN HOUSEHOLD%Total	PRESENCE AND TYPES OF INTERNET SUBSCRIPTIONS IN HOUSEHOLD%Total	PRESENCE AND TYPES OF INTERNET SUBSCRIPTIONS IN HOUSEHOLD%Total	PRESENCE AND TYPES OF INTERNET SUBSCRIPTIONS IN HOUSEHOLD%Total
								Total:	With an Internet subscription	Dial-up with no other type of Internet subscription	Broadband of any type	Cellular data plan	Cellular data plan with no other type of Internet subscription	Broadband such as cable, fiber optic or DSL	Broadband such as cable, fiber optic or DSL with no other type of Internet subscription	Satellite Internet service	Satellite Internet service with no other type of Internet subscription	Other service with no other type of Internet subscription	Internet access without a subscription	No Internet access

Supported Geographies (think “Sum Levels”)

1. List of supported geographies for the ACS endpoint:
<https://api.census.gov/data/2019/acs/acs5/geography.html>
2. We will also supply these to the API through Tidycensus.
(You can use the code or the keyword.)



Census Data API: FIPS Geographies in /data/2019/acs/acs5/geography

Reference Date	Geography Level	Geography Hierarchy
2019-01-01	010	us
2019-01-01	020	region
2019-01-01	030	division
2019-01-01	040	state
2019-01-01	050	state › county
2019-01-01	060	state › county › county subdivision
2019-01-01	067	state › county › county subdivision › subminor civil division
2019-01-01	070	state › county › county subdivision › place/remainder (or part)
2019-01-01	140	state › county › tract
2019-01-01	150	state › county › tract › block group
2019-01-01	155	state › place › county (or part)
2019-01-01	160	state › place
2019-01-01	170	state › consolidated city
2019-01-01	172	state › consolidated city › place (or part)

Sample API call with results:

1. Sample call: view the EXAMPLES page to see different possible combinations:
<https://api.census.gov/data/2019/acs/acs5/examples.html>
2. https://api.census.gov/data/2019/acs/acs5?get=NAME,B28002_005E&&for=county subdivision:*&in=state:42 & county:003&key=YOUR_KEY_HERE
3. Data is returned in JSON format (*JavaScript Object Notation*):

JSON	Raw Data	Headers
Save	Copy	Collapse All Expand All Filter JSON
▼ 0:		
0:	"NAME"	
1:	"B28002_005E"	
2:	"state"	
3:	"county"	
4:	"county subdivision"	
▼ 1:		
▼ 0:	"Bethel Park municipality, Allegheny County, Pennsylvania"	
1:	"9516"	
2:	"42"	
3:	"003"	
4:	"06064"	



JSON	Raw Data	Headers
Save	Copy	Pretty Print
[[{"NAME", "B28002_005E", "state", "county", "county subdivision",		
["Bethel Park municipality, Allegheny County, Pennsylvania", "9516", "42", "003", "06064"],		
["Bradford Woods borough, Allegheny County, Pennsylvania", "399", "42", "003", "08064"],		
["Monroeville municipality, Allegheny County, Pennsylvania", "8422", "42", "003", "50528"],		
["North Fayette township, Allegheny County, Pennsylvania", "4442", "42", "003", "55016"],		
["Ben Avon borough, Allegheny County, Pennsylvania", "604", "42", "003", "05504"],		
["Coraopolis borough, Allegheny County, Pennsylvania", "1682", "42", "003", "16144"],		
["Fawn township, Allegheny County, Pennsylvania", "591", "42", "003", "25400"],		
["Forest Hills borough, Allegheny County, Pennsylvania", "2187", "42", "003", "26592"],		
["Frazer township, Allegheny County, Pennsylvania", "289", "42", "003", "27608"],		
["Ingram borough, Allegheny County, Pennsylvania", "988", "42", "003", "37000"],		
["Kennedy township, Allegheny County, Pennsylvania", "2385", "42", "003", "39312"],		
["Lincoln borough, Allegheny County, Pennsylvania", "306", "42", "003", "43408"],		

So why use *Tidycensus*?



1. Tidycensus takes care of the formatting & API calls for you

“**tidycensus** is an R package that allows users to interface with the US Census Bureau's decennial Census and five-year American Community APIs and return *tidyverse-ready data* frames, optionally with simple feature (“SF”) geometry included. **tidycensus** is designed to help R users get Census data that is pre-prepared for exploration within the tidyverse, and optionally spatially with **sf**. “

2. What is the Tidyverse?



“The tidyverse is a collection of open source R packages introduced by Hadley Wickham and his team that ‘share an underlying design philosophy, grammar, and data structures’ of tidy data. “

Pt. 2: Dive into R & Tidycensus

Now for the fun part!



Let's move into R

1. Please open your RStudio Desktop or log into RStudio Cloud
2. Navigate to GitHub & copy the R code into a new script file:
 - https://github.com/cjtulley/PA_GeoDev_2022

Pt. 3: Resources + Q&A

Where you can learn more



Resources

1) Kyle Walker's website:

- eBook: "Analyzing US Census Data: Methods, Maps and Models in R"

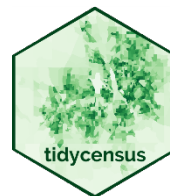
<https://walker-data.com/census-r/index.html>

- Tidycensus:

<https://walker-data.com/tidycensus/>

- Tigris:

<https://github.com/walkerke/tigris>



Kyle Walker: spatial data science
research & consulting

LinkedIn

Twitter

GitHub

Email

Website:

<https://walker-data.com/>

TIP: Join Walker's email list or follow him on social media to learn about his upcoming workshops and other updates.

Resources

2) U.S. Census Bureau Developer resources:

- Slack channel: <https://uscensusbureau.slack.com>



Developers' Forum

3) Developers landing page:

- <https://www.census.gov/data/developers.html>
 - Sign up for the Developers email list

Click here to join the Slack channel:

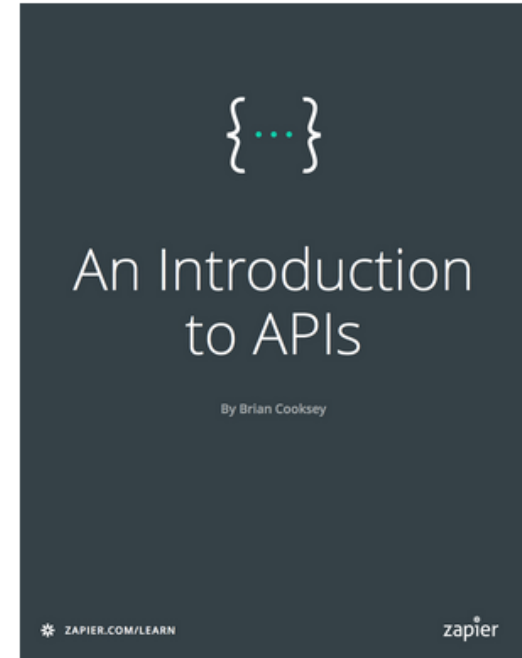
<https://bit.ly/3MDefaK>

TIP: Check the Slack channel for assistance - your fellow users are available to help you!

Resources

4) Zapier: An Introduction to APIs

- Free e-course: <https://zapier.com/learn/apis/>
- PDF available
- Table of Contents:
 - Chapter 1: Introduction
 - Chapter 2: Protocols
 - Chapter 3: Data Formats
 - Chapter 4: Authentication, Part 1
 - Chapter 5: Authentication, Part 2
 - Chapter 6: API Design
 - Chapter 7: Real-Time Communication
 - Chapter 8: Implementation



Thank you!



Regional Data Center

Southwestern Pennsylvania Commission
Data & Modeling

Catherine Tulley (she/her)

Transportation Planning Data Analyst

Southwestern Pennsylvania Commission

412-391-5590 | <https://www.spcregion.org>

ctulley@spcregion.org

GitHub: <https://github.com/cjtulley>