

Using the Census Bureau's API

With Python / Pandas Notebooks

Frank Donnelly

Newman Library, Baruch College CUNY

Spring 2020

Intro to the Census API

Understanding Datasets, Variables, and
Geography

Examples with Python and Jupyter Notebook

To Learn More

The US Census

It's *not* just the count that takes place every 10 years:

- ▶ Rich collection of several different datasets
- ▶ Summarizes people, houses, and businesses by place and subject
- ▶ Uses consistent definitions and terminology
- ▶ Geographically detailed
- ▶ Collected over varying periods of time

`https://data.census.gov/`



Explore Census Data

The Census Bureau is the leading source of quality data about the nation's people and economy.

 I'm looking for ...

[Advanced Search](#)



Choices and Filters

- ▶ Dataset
- ▶ Time frame
- ▶ Geography
- ▶ Topics / Subject

Then use filters to narrow your search

TOPICS **GEOGRAPHIES** YEARS SURVEYS INDUSTRIES

Q Enter text to filter...

« GEOGRAPHIES

- ☒ United States
- Region
- Division
- State
- County**

« COUNTY (STATE)

- Arizona
- Arkansas
- California**
- Colorado
- Connecticut
- Delaware

« CALIFORNIA

- ☒ All counties in California
- ☐ Alameda County, California
- ☐ Alpine County, California
- ☐ Amador County, California
- ☐ Butte County, California

REST API

REST = Representational State Transfer

API = Application Programming Interface

Allows end users to access services via the HTTP protocol using a predefined set of URLs that represent resources that you can interact with. By passing predefined variables into a links, you can access and retrieve data from that service within programs and scripts that you write.

Census Bureau's APIs

Main website: <https://www.census.gov/developers/>

Use it to:

- ▶ Avoid multiple clicks through graphic interfaces
- ▶ Create surgical extracts variable by variable
- ▶ Pull data into a language that you'll use for processing, visualization, and analysis
- ▶ Create well-documented and repeatable processes

Don't use for:

- ▶ Looking up simple facts
- ▶ Getting a pre-compiled table in one quick download
- ▶ Downloading massive amounts of data

Basic Example

`https://api.census.gov/data/`

```
import requests
popurl = 'https://api.census.gov/data/2018/pep/population?get=
        GEONAME,DATE_DESC,POP&for=county:061&in=state:36'
response=requests.get(popurl)
print(response.text)
```

```
[["GEONAME", "DATE_DESC", "POP", "state", "county"],
 ["New York County, New York",
  "7/1/2018 population estimate", "1628701", "36", "061"]]
```


Define Variables and Build Base URL

```
import requests

year='2018'
dsouce='pep'
dname='population'
cols='GEONAME,POP'
state='36'
county='005,047,061,081,085'
base_url = f'https://api.census.gov/data/{year}/{dsouce}/{dname}'

print(base_url)
```

'https://api.census.gov/data/2018/pep/population'

Pass in Variables and Get

```
data_url = f'{base_url}?get={cols}&for=county:{county}&in=state:{state}'  
print(data_url)
```

`https://api.census.gov/data/2018/pep/population?get=GEONAME,POP&for=county:005,047,061,081,085&in=state:36`

```
response=requests.get(data_url)  
print(response.text)
```

```
[["GEONAME", "POP", "state", "county"],  
["Bronx County, New York", "1432132", "36", "005"],  
["Kings County, New York", "2582830", "36", "047"],  
["New York County, New York", "1628701", "36", "061"],  
["Queens County, New York", "2278906", "36", "081"],  
["Richmond County, New York", "476179", "36", "085"]]
```

Today's Topics

Intro to the Census API

Understanding Datasets, Variables, and
Geography

Examples with Python and Jupyter Notebook

To Learn More

Census Datasets

Each dataset includes different variables and API URLs

Decennial census: 100% count of the population every ten years, includes basic demographic variables for all geographies

American Community Survey: sample survey of 3.5 million addresses captures detailed socio-economic characteristics for most geographies. Estimates with margins of error published as 1-year and 5-year averages; 1-year limited to places that have at least 65k people


Population estimates: annual estimates of basic demographic characteristics calculated using births, deaths, and migration for large geographies

County Business Patterns: counts of business establishments by industry and geography with data on employees and wages for large geographies

Start with the Available List

Census API Homepage:

`https://www.census.gov/developers/`




Search

BROWSE BY TOPICEXPLORE DATALIBRARYSURVEYS/ PROGRAMSINFORMATION FOR...FIND A CODEABOUT US

// Census.gov > Data > Developers > Available APIs

About this Section
API Forum
App Gallery
Available APIs
Geography
Guidance for Developers
News
Terms of Service
Updates
[Back to Developers](#)



Available APIs

We plan on adding more of our publicly available datasets. Here you'll find which of our many data sets are currently available via API. To make specific requests for the release of datasets, please sign up and submit your requests on our [Developer Forum](#).

NEW: We now have a machine-readable dataset discovery service available in beta release. Visit our [Discovery Tool](#) page to learn more.

Page 1 of 2 >

OCTOBER 17, 2019

American Community Survey 1-Year Data (2011-2018)

Areas with populations of 65,000+. Covers a broad range of topics about social, economic, demographic, and housing characteristics of the U.S. population.

JANUARY 30, 2020




American Community Survey 1-Year Supplemental Data (2014 - 2018)


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 (2012-2013)

Areas with populations of 20,000+. Covers a broad range of topics about social, economic, demographic, and housing characteristics of the U.S. population.

Is this page helpful? 
 Yes  No



Get Info About the Dataset

About this Section

API Forum

App Gallery

Available APIs

Geography

Guidance for Developers

News

Terms of Service

Updates

< Back to Population Estimates and Projections (2010 - 2019)



Population Estimates APIs



JANUARY 07, 2020



Each year, the Census Bureau's **Population Estimates Program** uses current data on births, deaths, and migration to calculate population change since the most recent decennial census and produces a time series of estimates of population, demographic components of change, and housing units. The annual time series of estimates begins with the most recent decennial census data and extends to the vintage year. As each vintage of estimates includes all years since the most recent decennial census, the latest vintage of data available supersedes all previously-produced estimates for those dates.

Vintage 2019	Vintage 2018	Vintage 2017	Vintage 2016	Vintage 2015	Vintage 2014	Vintage 2013	2000-2010 Intercensals	MORE ▾
--------------	---------------------	--------------	--------------	--------------	--------------	--------------	------------------------	--------

Vintage 2018

Note: These products were redesigned in 2015 to provide a better user experience for developers.

Population Estimates

- API Call: api.census.gov/data/2018/pep/population
- Examples and Supported Geographies: api.census.gov/data/2018/pep/population.html
- Variables: api.census.gov/data/2018/pep/population/variables.html

Related Information

[Population Estimates Categorical Variables](#)

You May Be Interested In

MOST POPULAR
[2020CENSUS.GOV](https://2020census.gov)



Is this page helpful?

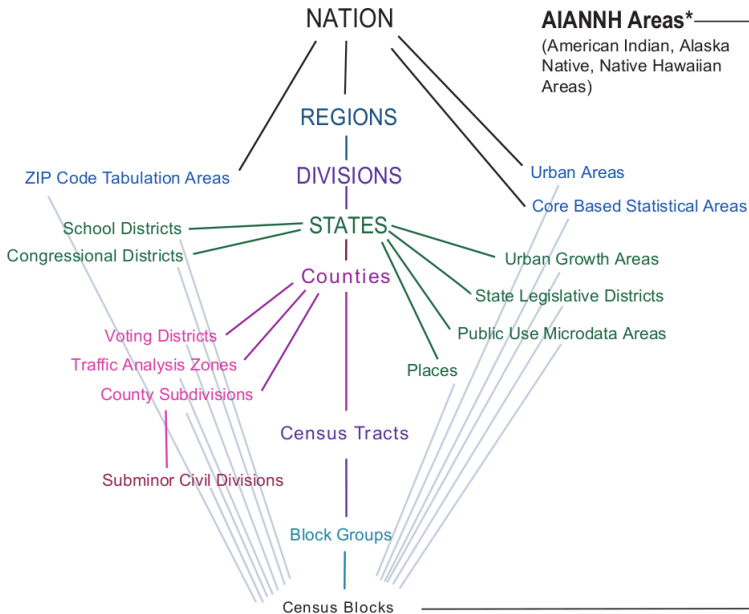
Yes No

Inspect the Variables

Census Data API: Variables in /data/2018/pep/population/variables

Name	Label	Concept	Required	Attributes	Limit	Predicate Type	Group
CBSA	Metropolitan Statistical Area/Micropolitan Statistical Area code		not required		0	int	N/A
CONCIT	Consolidated city FIPS code		not required		0	int	N/A
COUNTY	County FIPS code		not required		0	int	N/A
COUSUB	Minor Civil Division FIPS code		not required		0	int	N/A
CSA	Combined Statistical Area code		not required		0	int	N/A
DATE_CODE	Date		default displayed		0	int	N/A
DATE_DESC	Description of DATE values		not required		0	string	N/A
DENSITY	Population density as of July 1 of the vintage year only		not required		0	int	N/A
DIVISION	Census Division Code		not required		0	(not a predicate)	N/A
for	Census API FIPS 'for' clause	Census API Geography Specification	predicate-only		0	fips-for	N/A
FUNCSTAT	Functional Status Code		not required		0	string	N/A
GEO_ID	Geographic identifier code		not required		0	string	N/A
GEONAME	Geographic Name		not required		0	(not a predicate)	N/A
in	Census API FIPS 'in' clause	Census API Geography Specification	predicate-only		0	fips-in	N/A
LASTUPDATE	Last Update		not required		0	string	N/A
MDIV	Metropolitan Division code		not required		0	int	N/A
NATION	Nation		not required		0	(not a predicate)	N/A
PLACE	Place FIPS code		not required		0	int	N/A
POP	Population		not required		0	(not a predicate)	N/A
PRIMGEOFLAG	Primitive Geography Flag(1=Yes; 0=No)		not required		0	int	N/A
REGION	Census Regional Code		not required		0	(not a predicate)	N/A
STATE	State FIPS code		not required		0	(not a predicate)	N/A
SUMLEV	Summary Level		not required		0	string	N/A
ucgid	Uniform Census Geography Identifier clause	Census API Geography Specification	predicate-only		0	ucgid	N/A
UNIVERSE	Universe		default displayed		0	string	N/A
25 variables							

Census Geography



Check the Geographies

Census Data API: FIPS Geographies in /data/2018/pep/population/geography

Reference Date	Geography Level	Geography Hierarchy	Limit
2018-01-01	01	us	
2018-01-01	02	region	
2018-01-01	03	division	
2018-01-01	04	state	
2018-01-01	050	state › county	
2018-01-01	061	state › county › county subdivision	
2018-01-01	071	state › county › county subdivision › place	
2018-01-01	157	state › county › place	
2018-01-01	162	state › place	
2018-01-01	170	state › consolidated city	
2018-01-01	172	state › consolidated city › place	
2018-01-01	310	metropolitan statistical area/micropolitan statistical area	
2018-01-01	314	metropolitan statistical area/micropolitan statistical area › metropolitan division	
2018-01-01	330	combined statistical area	
14 items			

Geographic Identifiers

Geographies are uniquely identified using ANSI / FIPS codes:
<https://census.missouri.edu/geocodes/>

State-County (34 New Jersey, 005 Burlington County): 34005

State-County (36 New York, 005 Bronx County): 36005

State-County-Tract (36 New York, 005 Bronx County, Tract 1): 36005000100

State-Place (36 New York, 51000 New York City): 3651000

Zip Code Tabulation Area (ZCTA 10010): 10010

Geographies For and In

```
#Get Kings and Queens County in New York State  
&for=county:047,081&in=state:36
```

```
#Get all counties in New York State  
&for=county:*&in=state:36
```

```
#Get New York State  
&for=state:36
```

```
#Get New Jersey and New York  
&for=state:34,36
```

```
#Illegal call, as you must specify a state to get counties  
&for=county:047,081
```

Today's Topics

Intro to the Census API

Understanding Datasets, Variables, and
Geography

Examples with Python and Jupyter Notebook

To Learn More

You can access the Census API using many different languages and environments. We're using Python in a Jupyter Notebook in these examples.

- ▶ Anacondas: a Python distribution with many prepackaged modules and tools
 - ▶ Python: a powerful, flexible, general purpose programming language
 - ▶ Requests: a module for using APIs
 - ▶ Pandas: a module for data science and viz
 - ▶ Jupyter Notebook: web-based programming environment that lets you run iterative and annotated blocks of codes
 - ▶ Spyder: traditional programming environment with variable and file explorer

Python / Pandas / Notebook tips: *Practical Business Python*
<https://pbpython.com/>

Examples

Download from:

https://github.com/frankpd/census_api_examples

Example A : Hardcoded variables, basic retrieval, Pandas operations (Population Estimates)

Example B : Using an API key, saving retrieved data (Population Estimates)

Example C : Reading variables from files, looping (ZIP Code Business Patterns)

Example A

- ▶ Save all hardcoded variables at the top of your program
- ▶ Save attributes and geographies as single strings with values separated by commas
- ▶ Requests module for using APIs
- ▶ Build a base url and a data url
- ▶ Geographies must follow census hierarchy
- ▶ Retrieved data includes column headers and geography values from your URL
- ▶ Metadata for variables can be captured separately

Example B

- ▶ You are limited to 500 requests per day without an API Key
- ▶ Register to get an API key:
`https://www.census.gov/developers/`
- ▶ You tack your key variable to the end of a url when making requests
- ▶ Do not hardcode your key value into your script
- ▶ Store key in a text file, read the file in, save as variable
- ▶ If using git / GitHub, create a gitignore file to ignore the key file
- ▶ Once you successfully retrieve data, dump to a file
- ▶ Read data from file in the next block, to avoid having to call the API again

Example C

- ▶ If you have many attributes or geographies to retrieve, read them in from files
- ▶ You cannot insert more than 50 variables into a single request
- ▶ Requesting attributes or geographies one at a time can be slow, not feasible for big requests
- ▶ For big requests, create a list with sublists of 50 variables
- ▶ Loop through these sublists, calling from the API and saving as you go
- ▶ Column headers are returned with each request, you only want them once
- ▶ Refamiliarize yourself with basic Python data structures, lists and dictionaries

Python Data Structures

List: Each sublist is a record, column meanings are implicit, manipulate by looping and slicing of indices

```
[[ 'New Jersey', 8.9], [ 'New York', 19.5]]
```

Dictionary: series of key - value pairs where key is unique, iterate through keys, values, or items. Can also nest

```
{ 'New Jersey': 8.9, 'New York': 19.5 }
```

Hybrid: nest lists in dictionaries or vice versa and use methods and functions of both

```
{34: [ 'New Jersey', 8.9], 36:[ 'New York', 19.5]}
```

Dataframe: with Pandas, a gridded structure with explicit rows and columns

```
({ 'Name' : [ 'New Jersey', 'New York'], 'Population' :  
[8.9,19.5]}, index = [ '34', '36'] )
```

Today's Topics

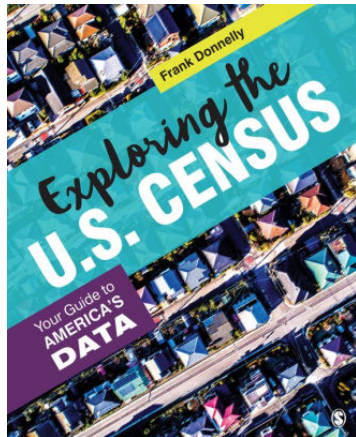
Intro to the Census API

Understanding Datasets, Variables, and
Geography

Examples with Python and Jupyter Notebook

To Learn More

Exploring the US Census: Your Guide to America's Data
SAGE Publications, 2020



<https://atcoordinates.info/census-book/>

Contact

Frank Donnelly, Geospatial Data Librarian, Baruch CUNY
francis.donnelly@baruch.cuny.edu | 646-312-1657



<http://guides.newman.baruch.cuny.edu/gis>