

## Instructions for accessing 2019 PUMS data in Python

If you already have Python and its associated packages installed, please skip to step 3.

1. Download and install Python
  - Go to <https://www.python.org/downloads/> and download Python then install it
2. Install necessary packages
  - Pandas
    - Windows: in command line run `pip install pandas`
    - Mac: in terminal run `conda install pandas`
  - Requests
    - Windows: in command line run `pip install requests`
    - Mac: in terminal run `conda install requests`
  -
3. You will need to get an API key from the Census.
  - Go to [https://api.census.gov/data/key\\_signup.html](https://api.census.gov/data/key_signup.html)
  - Fill out form
    - Organization Name: **\*\*School Name\*\***
    - Email address: **\*\*School email address (or personal)\*\***
    - Agree to terms of services
    - Click "Submit Key request"
4. You will receive an email telling you to activate your API key.
  - Click the link to activate your key.
  - The email will also give you your API key
5. Start running Python commands with import statements.
  - Open Python Idle
  - Enter the following commands:
    - `import pandas as pd`
    - `import numpy as np`
    - `import requests`
    - `import json`
6. Copy the API key you received from step 4
7. Prepare for the API request by making a variable for your API key and setting the variables you wish to retrieve. Note: this is only a few variables, you will need to dictate in your command which variables you would like to get. Please run the following commands:
  - `api_key='**your API key**'`
  - `variables='PINCP,WAGP,INTP'`
  - `url='https://api.census.gov/data/2019/acs/acs1/pums?get='+variables+'&key='+api_key`
8. Make the API request and check its status:
  - `response=requests.get(url)`
  - `print(response.status_code)`
  - A list of possible response codes and what they indicate:

- 1xx (Informational): Communicates transfer protocol-level information
- 2xx (Success): Client's request was accepted successfully (what we want)
- 3xx (Redirection): Client must take some additional action in order to complete their request
- 4xx (Client Error): Points the finger at the client (check your code!)
- 5xx (Server Error): Server takes responsibility for error

9. Next, we will load and make the data readable then print it.

- `census = response.text`
- `census = json.loads(census)`
- `df = pd.DataFrame.from_dict(census)`
- `print(df)`
- This gets a response such as

```

0          0          1          2
0      PINCP      WAGP      INTP
1      63000      63000          0
2          0          0          0
3      2000      2000          0
4          0          0          0
...      ...      ...      ...
3239549 150000 150000          0
3239550  27000          0          0
3239551  75000      75000          0
3239552  57000      57000          0
3239553  55000      55000          0
[3239554 rows x 3 columns]
```

10. To make the data more intuitive, we'll reset the header and give the columns names.

- `new_header = df.iloc[0]`
- `df = df[1:]`
- `df.columns = new_header`
- `df.columns = [`  
`'INCOME',`  
`'SALARY',`  
`'INTEREST_DIV'`  
`]`
- `print(df.head())`
- This now gets us the following

```

      INCOME  SALARY  INTEREST_DIV
1    63000    63000             0
2         0         0             0
3     2000     2000             0
4         0         0             0
5     6000     6000             0
```

11. Now you have the basics for pulling the data from the census API and can use these steps to pull different variables and begin your own analysis.

These instructions are inspired and modified from <https://towardsdatascience.com/using-the-us-census-api-for-data-analysis-a-beginners-guide-98063791785c>. Please go to their website for more details and further instructions if desired.