Jupyter: Intro to Data Science - Lecture 2 Loading and Summarizing Data

Grant Long

CUNY City College

NYC Tech-in-Residence Corps

Follow this and additional works at: https://academicworks.cuny.edu/cc_oers

Part of the Computer Sciences Commons

How does access to this work benefit you? Let us know!

Recommended Citation

This Lecture or Presentation is brought to you for free and open access by the City College of New York at CUNY Academic Works. It has been accepted for inclusion in Open Educational Resources by an authorized administrator of CUNY Academic Works. For more information, please contact AcademicWorks@cuny.edu.
Data Dive 2: Loading and Summarizing Data

Part 1: Tapping the Census API for Valuable Insights

The U.S. Census API is an extremely rich data source by itself, but also one that can enhance and provide insights into other datasets. In the first part of today’s exercise, we’ll put

Helpful links:

- API user guide (https://www.census.gov/data/developers/guidance/api-user-guide.html)
- List (https://api.census.gov/data/2016/acs/acs5/variables.html) of census variables provided by the API.
- Sample queries (https://api.census.gov/data/2016/acs/acs5/examples.html) for different geographies
- List (https://api.census.gov/data.html) of available apis
- The Census Factfinder (https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml) is a great place to browse data or look up variable names.

Today we’ll be using the 2016 ACS 5-year sample. The details of the different ACS samples are beyond the context of this class, but suffice to say that these estimates are built of five years of data from the ongoing ACS.

The Census API, like many others, requires a developer key to use the API regularly. You can sign up for one here (https://api.census.gov/data/key_signup.html). Thus far, I have not needed one for these exercises, but just to be on the safe side, it’s nice to have one handy.

```python
In [ ]:
import pandas as pd
import requests

# In case I need a key, I'll leave these here.
import os

census_key = os.getenv('CENSUS_KEY')
```

Using the API

Some of the many useful variables in the ACS:

- B01001_001E: Total Population
- B19013_001E: Median Household Income
- B25058_001E: Median Rent
- B01002_001E: Median Age

We’ll need the base url for the endpoint, in this case https://api.census.gov/data/2016/acs/acs5?, along with the relevant parameters in our query. The requests package has a handy feature that allows you to enter your parameters as a dictionary using the params argument.
Load our results into a data frame. In this case, we set the first row of results as the headers.

Load our results into a data frame. In this case, we set the first row of results as the headers.

Let's clean things up a bit by renaming our columns and resetting the index

Now we can start to have some fun by digging through the data.

Writing to File

Pandas also has simple functionality to write to files, not just read from them. To write to a csv file to store our data, we can just use `.to_csv()`.

```python
# census_df.to_csv('..\..\..\web\www\data\census_counties.csv')
```
Exercise: America's Oldest Counties

In [ ]: # todo: find america's oldest counties by median age
In [ ]:
In [ ]:
In [ ]:
In [ ]: