Using Census APIs for Equity Analysis

Presentation to the 2023 TRB Annual Meeting

8 January 2023

Maddalena Romano, mromano1@dot.nyc.gov

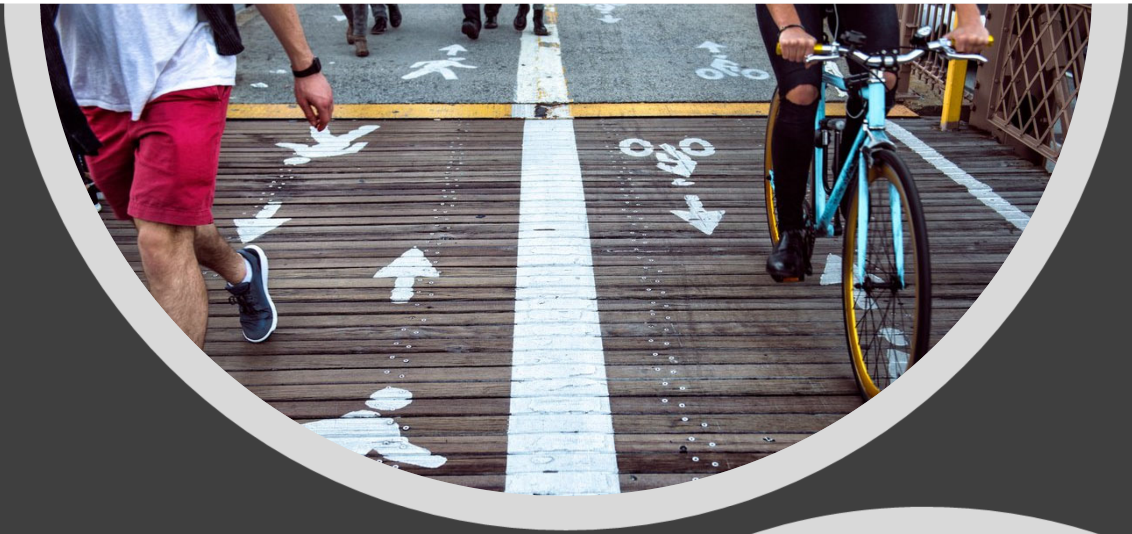
What is Equity?

Horizontal Equity

- Fair share
- External costs

Vertical Equity

- Inclusivity
- Affordability
- Social Justice



What To Do First?

Prerequisites

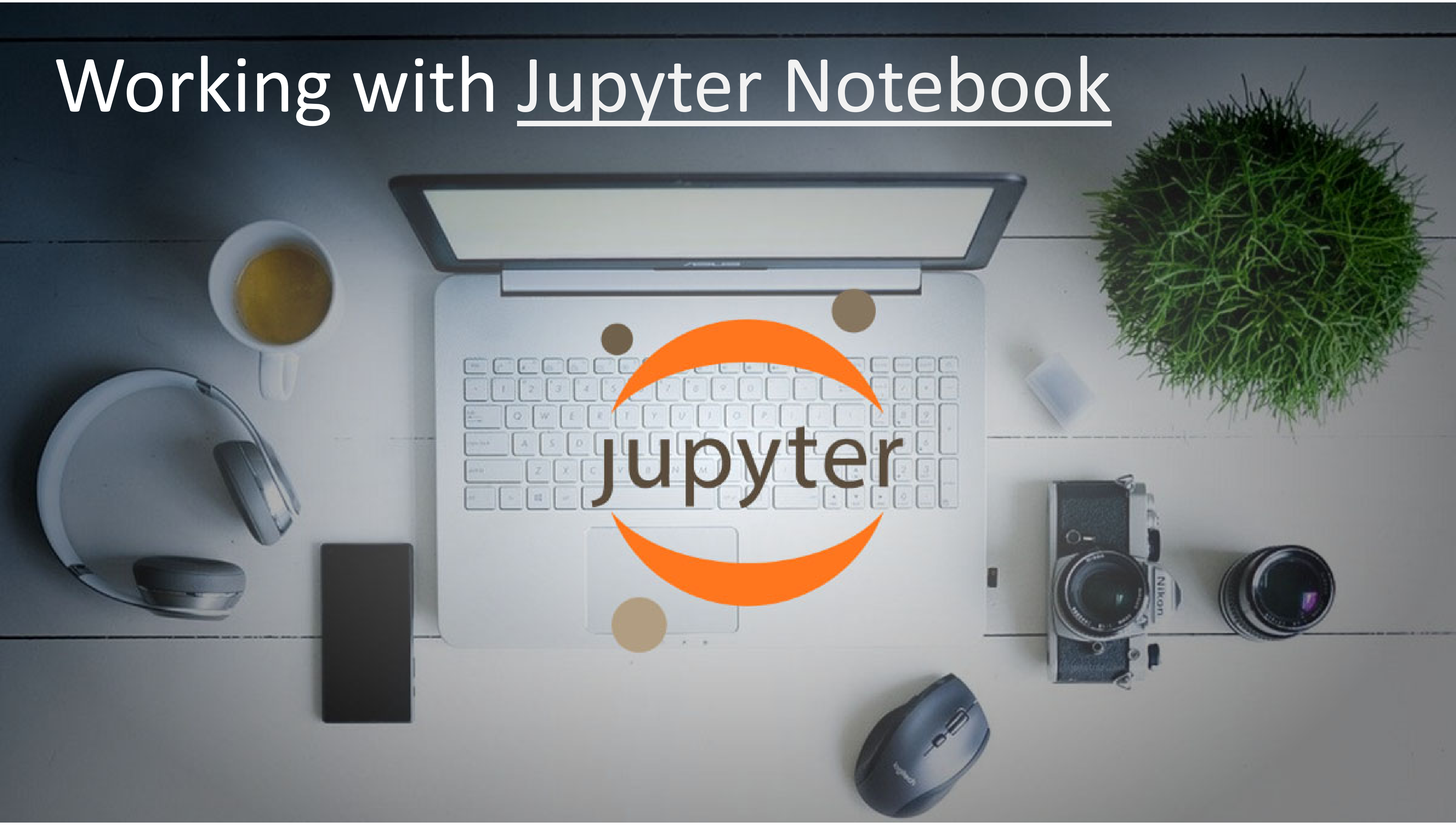
API Key

Locate Census tables you want

PREREQUISITES

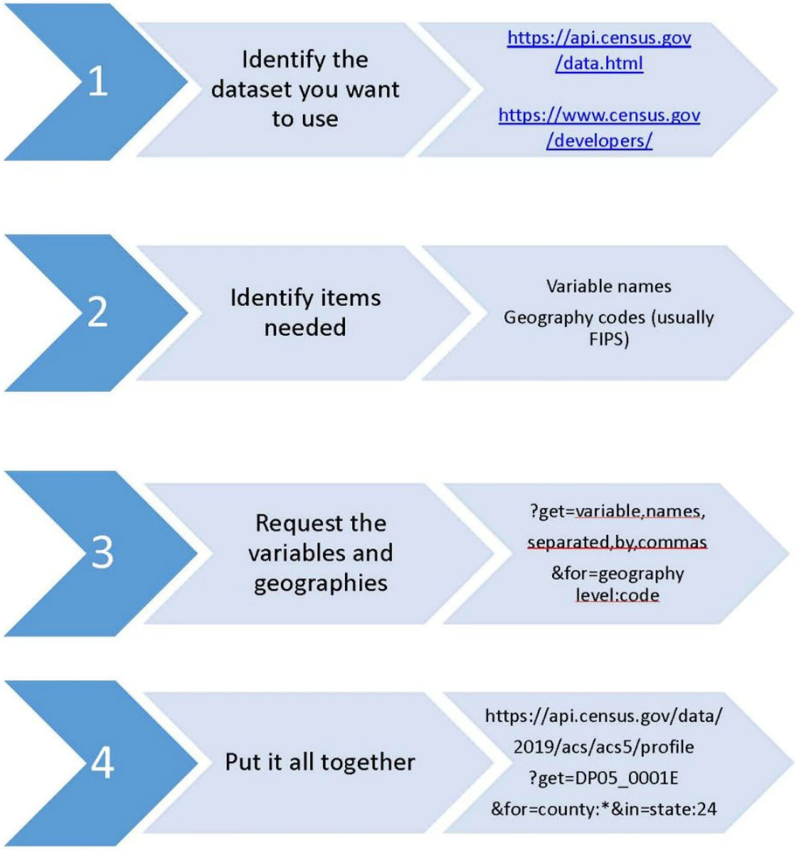
- Anaconda 3 installed
  - If needed, update shapely “conda install shapely”
- Knowledge of Jupyter Notebook and Python
- Knowledge of how to install GeoPandas. If that doesn’t work:
  - Type in “conda create –name py39 python=3.9” in the Anaconda Prompt and hit enter. Py39 is the virtual environment name
  - If that works, type in “activate py39”, press Enter
  - Type in “conda install geopandas” and hit enter
- Knowledge of how to perform pip installations
  - # pip install PyGithub
  - # pip install census
  - # pip install us
  - # pip install topojson

Working with Jupyter Notebook



Using the API

Source: https://data.census.gov

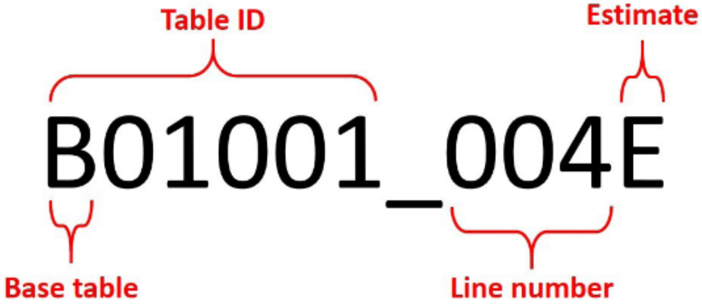


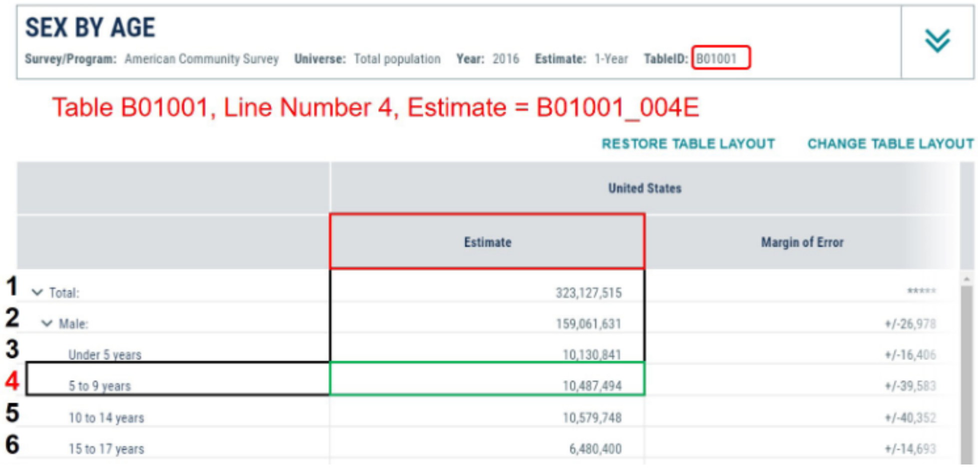
1. Identify the dataset you want to use — https://api.census.gov/data.html https://www.census.gov/developers/

2. Identify items needed — Variable names Geography codes (usually FIPS)

3. Request the variables and geographies — ?get=variable,names,separated,by,commas &for=geography level:code

4. Put it all together — https://api.census.gov/data/2019/acs/acs5/profile?get=DP05_0001E&for=county:*&in=state:24





SEX BY AGE

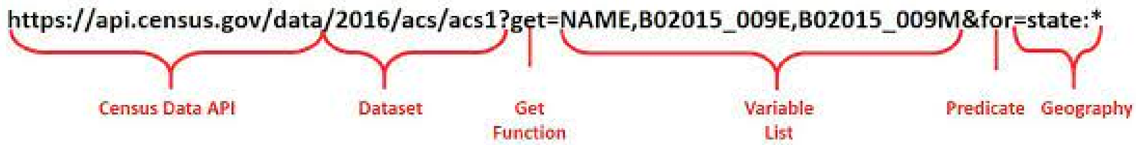
Survey/Program: American Community Survey Universe: Total population Year: 2016 Estimate: 1-Year TableID: B01001

Table B01001, Line Number 4, Estimate = B01001_004E

RESTORE TABLE LAYOUT CHANGE TABLE LAYOUT

		United States	
		Estimate	Margin of Error
1	Total:	323,127,515	*****
2	Male:	159,061,631	+/-26,978
3	Under 5 years	10,130,841	+/-16,406
4	5 to 9 years	10,487,494	+/-39,583
5	10 to 14 years	10,579,748	+/-40,352
6	15 to 17 years	6,480,400	+/-14,693

Base request Parameters



https://api.census.gov/data/2016/acs/acs1?get=NAME,B02015_009E,B02015_009M&for=state:*

What Next?

Start Coding

Visualize Data

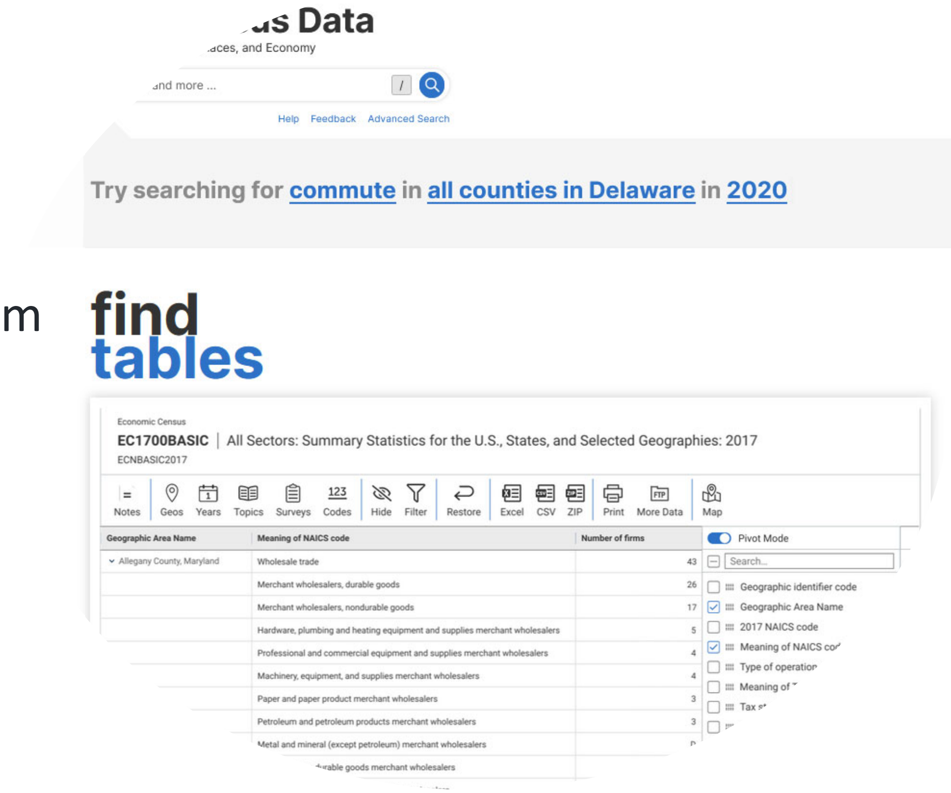
Export Data

IMPORT MODULES

```
import matplotlib.pyplot as plt # plotting tool
import pandas as pd # data analysis and manipulation tool
import geopandas as gpd # extends pandas to allow spatial operations on geometric types
from shapely.geometry import Polygon # for geometric operations
from census import Census # library for accessing census tables
from us import states # library for accessing the FIPS codes for many geographies
import fiona # for file access
import base64 # provides functions for encoding/decoding binary data to printable ASCII
from github.MainClass import Github, GithubIntegration # main class to access Github API
from github.InputGitTreeElement import InputGitTreeElement # i.e., path, type, shape
import os #  provides functions for interacting with the underlying operating system
```

Import data from Census

- Go to Census API key site and fill out webform
- Register your 40-digit API key code
- Access tables from list of variables in ACS
- Assign data to a variable
- Create dataframe
- Read data into dataframe using Pandas



Try searching for commute in all counties in Delaware in 2020

Economic Census

EC1700BASIC | All Sectors: Summary Statistics for the U.S., States, and Selected Geographies: 2017

ECNBASIC2017

Geographic Area Name	Meaning of NAICS code	Number of firms
Allegany County, Maryland	Wholesale trade	43
	Merchant wholesalers, durable goods	26
	Merchant wholesalers, nondurable goods	17
	Hardware, plumbing and heating equipment and supplies merchant wholesalers	5
	Professional and commercial equipment and supplies merchant wholesalers	4
	Machinery, equipment, and supplies merchant wholesalers	4
	Paper and paper product merchant wholesalers	3
	Petroleum and petroleum products merchant wholesalers	3
	Metal and mineral (except petroleum) merchant wholesalers	D
	rable goods merchant wholesalers	

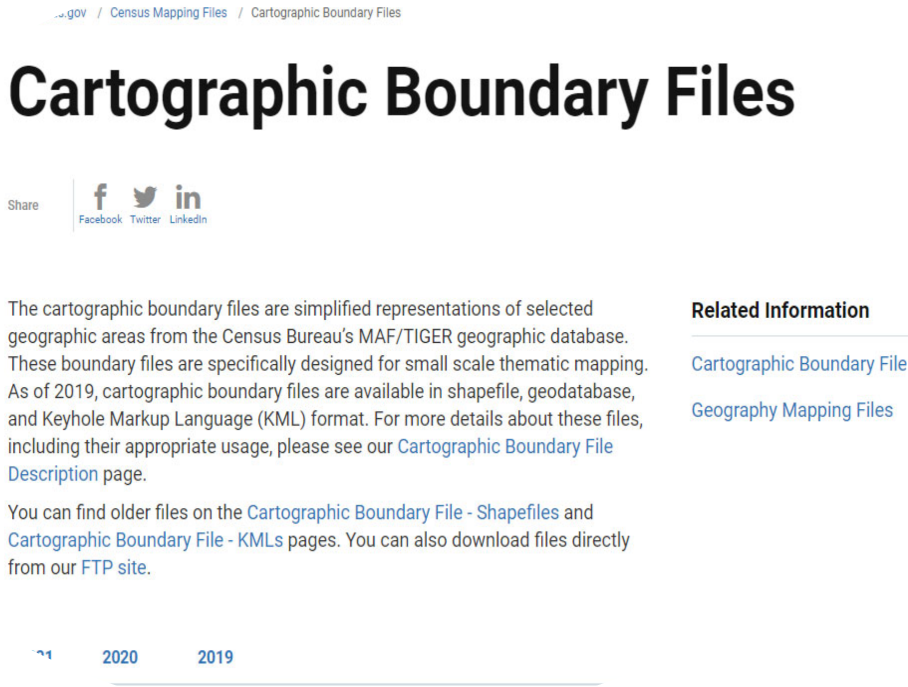
□ Geographic identifier code
☑ Geographic Area Name
□ 2017 NAICS code
☑ Meaning of NAICS co
□ Type of operation
□ Meaning of
□ Tax s

NOTE: Keep your API key private!

For more information, see Census API User Guide

Import Shapefile

- Download data from Cartographic Boundary Files page or the TIGER/Line Shapefiles page
- Reproject the spatial data into the UTM Zone 18N
- Create a geodataframe to hold this data
- Confirm that table and spatial data have the same number of rows



.gov / Census Mapping Files / Cartographic Boundary Files

Cartographic Boundary Files

The cartographic boundary files are simplified representations of selected geographic areas from the Census Bureau's MAF/TIGER geographic database. These boundary files are specifically designed for small scale thematic mapping. As of 2019, cartographic boundary files are available in shapefile, geodatabase, and Keyhole Markup Language (KML) format. For more details about these files, including their appropriate usage, please see our Cartographic Boundary File Description page.

You can find older files on the Cartographic Boundary File - Shapefiles and Cartographic Boundary File - KMLs pages. You can also download files directly from our FTP site.

Related Information

Cartographic Boundary File

Geography Mapping Files

Create new column and check data types

- Identify column with similar unique data
- If it doesn’t exist, create one using [] and the column name

```
ny_df["GEOID"] = ny_df["state"] + ny_df["county"] + ny_df["tract"]
```

- Then remove any columns no longer needed

```
ny_df = ny_df.drop(columns = ["state", "county", "tract"])
```

- And check data types for the census data and shapefile

```
print("Column data types for census data:\n{}".format(ny_df.dtypes))
print("\nColumn data types for census shapefile:\n{}".format(ny_tract.dtypes))
```

Note: Read about indexing and selecting data in the pandas documentation

Note: For more information, see the pandas help documentation on drop

Merge dataframes and make a subset

- Use GeoPandas to merge the two dataframes together

```
ny_merge = ny_tract.merge(ny_df, on = "GEOID")
print('Shape: ', ny_merge.shape)
```

- Clean up the dataframe of any unnecessary columns

```
ny_poverty_tract = ny_merge[["STATEFP", "COUNTYFP", "TRACTCE", "GEOID", "geometry",
"C17002_001E", "C17002_002E", "C17002_003E", "B01003_001E"]]
```

Note: The columns for the key do not need to have the same name, and the relationship doesn't need to be one-to-one. See this Esri ArcGIS help documentation on joins and relates for more information.

Dissolve geometries and perform math

- Use dissolve function to aggregate the poverty and population values for those tracts within the same county.

```
ny_poverty_county = ny_poverty_tract.dissolve(by = 'COUNTYFP', aggfunc = 'sum')
print(ny_poverty_county.head(5))
print('Shape: ', ny_poverty_county.shape)
```

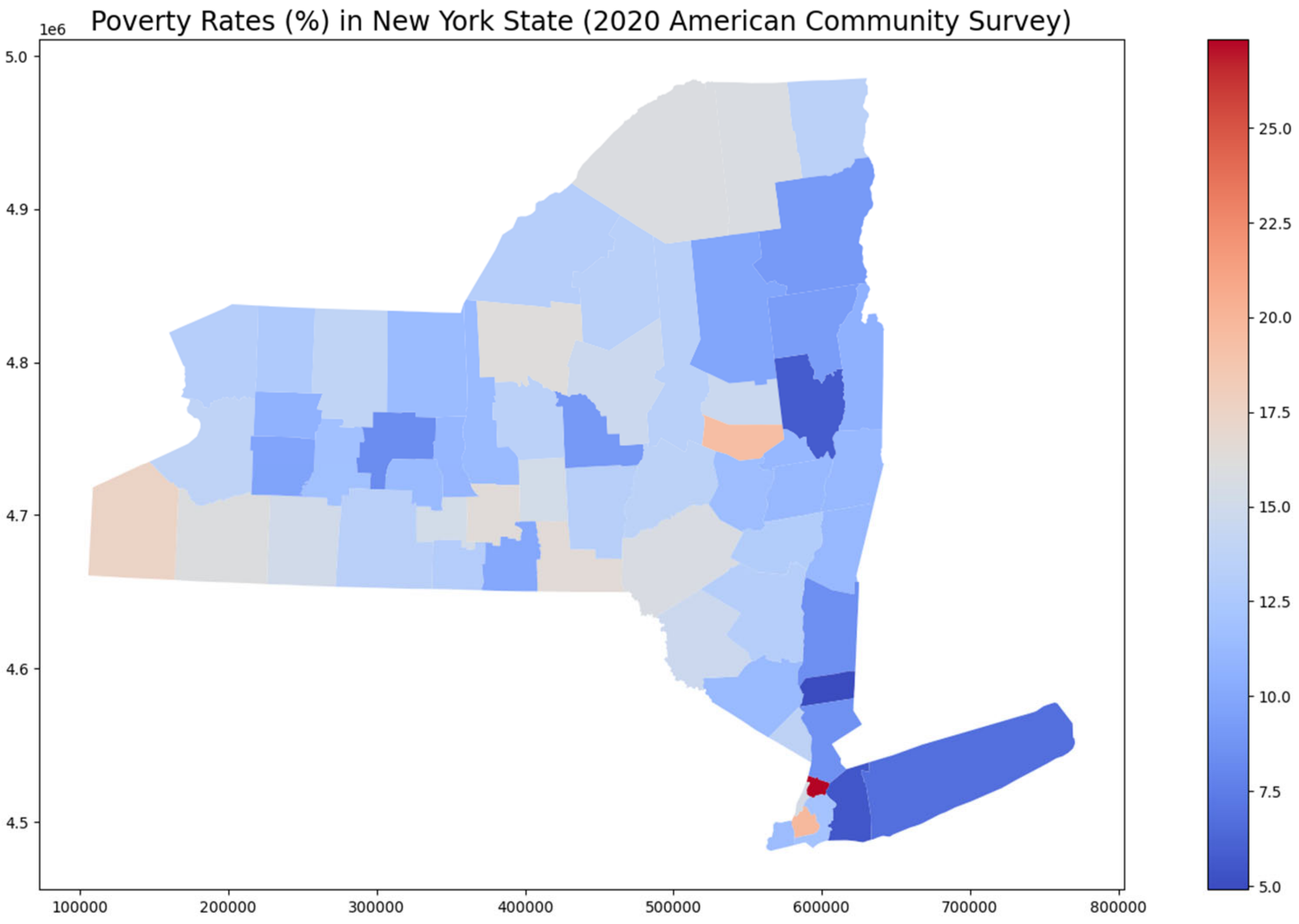
Note: Number of rows decreased from 4918 to 62.

- Estimate the poverty rate

```
ny_poverty_county["Poverty_Rate"] = (ny_poverty_county["C17002_002E"] +
ny_poverty_county["C17002_003E"]) / ny_poverty_county["B01003_001E"] * 100
ny_poverty_county.head(5)
```

Plotting results

- Create subplot
- Plot data
- Add styling
- Add title



Note: Check this list for alternative color palettes

Write data to a file

- Create output path to store your data
- Set up error handling
- Write file to directory
  - Set output type

```
try:
   ny_poverty_county.to_file(r"\\insert\your\directory\here.shp", encoding='utf-8')
   print("Shapefile successfully written to directory")
except OSError as error:
   print ("Shapefile cannot be written to directory")
```

THANK YOU!

Questions?

Scan me

Maddalena Romano, mromano1@dot.nyc.gov

https://github.com/mromano1/Using-Census-API-for-Equity-Analysis

NEW YORK CITY DOT

NYC DOT NYC DOT nyc_dot NYC DOT