

Final Proposal: The Dynamics of Economic Activity at the Community Level.

Principal Investigator: Michael Waugh (<https://www.waugh-economics.com>)
Email: mwaugh@stern.nyu.edu (mwaugh@stern.nyu.edu)

This project will study how measures of economic activity (income, unemployment, production structure) evolve over time at a narrow geographic level (a [commute zone](https://en.wikipedia.org/wiki/Commuting_zone)) within the United States. As an example, consider New York City: It has had ups and downs over the years. In the 1970s and 80s it was a model of urban decay; Today, there are few areas in the United States that are more prosperous than New York. This project will describe and visualize the properties of these fluctuations.

The key element of the project is the use of [BEA's API](https://www.bea.gov/API/bee_web_service_api_user_guide.htm) providing access to measures of economic activity at (i) detailed geographic levels (e.g. county's) and (ii) a long time series for these measures dating back to 1969. Details of this dataset are described below in the data report.

I anticipate that the project will have three sections.

- Basic statistics about the cross-sectional inequality and the volatility of income over time will be reported.
- Visualizations of illustrating the dynamics. For example, paths for interesting areas will be presented relative to aggregate statistics of the distribution, e.g. like Korea in the PWT matplotlib example) and disasters (e.g. Detroit).
- Finally, I plan present a national map that illustrates the relative changes over time to visualize the geographic distribution of these local economic fluctuations in economic activity. This map might be supplemented with an additional map which cross-references the production structure of each community, e.g. we may see that those communities that were most tilted towards manufacturing are the ones that declined the most.

Data Report

Overview: The data behind my project comes from the [Bureau of Economic Analysis](https://www.bea.gov/). As mentioned above, their [regional accounts data](https://www.bea.gov/regional/index.htm) provides access to measures of economic activity at (i) detailed geographic levels (e.g. county's) and (ii) a long time series for these measures dating back to 1969.

Important Variables: The key series that I must retrieve is what the BEA calls [personal income](https://www.bea.gov/newsreleases/regional/lapi/lapi_newsrelease.htm) which is defined as:

"*Personal income* is the income received by, or on behalf of, all persons from all sources: from participation as laborers in production, from owning a home or unincorporated business, from the ownership of financial assets, and from government and business in the form of transfer receipts. It includes income from domestic sources as well as from the rest of the world."

One way to think about this is that it is close to an income side measure of output. In my analysis, I will focus on personal income per capita which adjusts this measure by the population within that geographic area.

In my report I will also download population. Other measures of economic activity are available from the BEA or can be merged with the Census.

The *Geography* that I will work with is at the county level. I eventually may aggregate up to what is called a "commute zone."

Access I will use the BEA's API to download and access the data. Below I demonstrate that I have the ability to access the data.

Requisite Packages Below I bring in the packages I need...

```
In [2]: import pandas as pd # We know this one...
import requests # This is usefull with the API
import numpy as np # For performing numerical analysis
import matplotlib.pyplot as plt # Plotting
import weightedcalcs as wc # This allows for "weighted" calculations
```

Grabing the Data: Below I use the BEA API to grab personal income per capita and population for all years between 1969 and 2015. First, I create a string of years to pass to the BEA.

```
In [4]: years = range(1969,2018)

years = "".join(str(list(years)))

years = years[1:-1]
```

Then I will grab the income data...

```
In [5]: BEA_ID = "6BF79D8C-8042-4196-88DC-0E0C55B0C3B6" # This is my Key

my_key = "https://bea.gov/api/data?&UserID=" + BEA_ID + "&method=GetData&"

data_set = "datasetname=RegionalIncome&" # This access the Regional Income dataset

table_and_line_income = "TableName=CA1&LineCode=3&" # This grabs the income data

table_and_line_population = "TableName=CA1&LineCode=2&" # This grabs the populaiton data

year = "Year=" + years + "&" # Makes the years

location = "GeoFips=COUNTY&" # This is the location. I'm going to do this at the county level.

form = "ResultFormat=json" # The format.
```

```
In [6]: API_URL = my_key + data_set + table_and_line_income + year + location + form

r = requests.get(API_URL)

df_income = pd.DataFrame(r.json()["BEAAPI"]["Results"]["Data"])
```

Then I'm going to clean this up a bit...

```
In [ ]: df_income.drop(['CL_UNIT', 'Code', "NoteRef", "UNIT_MULT"], axis=1, inplace = True)

#df["DataValue"].column = "IncomePC"

df_income.rename(columns={"DataValue": "IncomePC"}, inplace=True)
```

Then do the same thing for population...

```
In [ ]: API_URL = my_key + data_set + table_and_line_population + year + location + form

r = requests.get(API_URL)

population = pd.DataFrame(r.json()["BEAAPI"]["Results"]["Data"])

population.drop(['CL_UNIT', 'Code', "NoteRef", "UNIT_MULT", "GeoName"], axis=1, inplace = True)

#df["DataValue"].column = "IncomePC"

population.rename(columns={"DataValue": "Population"}, inplace=True)

population.head()
```

Then merge the two datasets together...

```
In [17]: combo = pd.merge(population, df_income, # Left df, right df
                        how='inner', # Try the different options, inner, outer, left, right...what happens.
                        on=['GeoFips', "TimePeriod"], # Link with cntry
                        indicator=True) # Tells us what happend
```

These are some summary statistics from about the data

In [21]: `combo.info()`

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 153504 entries, 0 to 153503
Data columns (total 6 columns):
Population    153504 non-null object
GeoFips       153504 non-null object
TimePeriod    153504 non-null object
IncomePC      153504 non-null object
GeoName       153504 non-null object
_merge        153504 non-null category
dtypes: category(1), object(5)
memory usage: 7.2+ MB
```

So there is a decent amount of stuff here. Let me show you an aspect of one year by looking at the head and the tail.

In [23]: `combo[combo.TimePeriod == "2015"].head(10)`

Out[23]:

| | Population | GeoFips | TimePeriod | IncomePC | GeoName | _merge |
|-----|------------|---------|------------|----------|---------------|--------|
| 4 | 320896618 | 00000 | 2015 | 48451 | United States | both |
| 54 | 4853875 | 01000 | 2015 | 38214 | Alabama | both |
| 100 | 55035 | 01001 | 2015 | 38575 | Autauga, AL | both |
| 150 | 203690 | 01003 | 2015 | 40640 | Baldwin, AL | both |
| 196 | 26270 | 01005 | 2015 | 31635 | Barbour, AL | both |
| 246 | 22561 | 01007 | 2015 | 28919 | Bibb, AL | both |
| 292 | 57676 | 01009 | 2015 | 31560 | Blount, AL | both |
| 342 | 10455 | 01011 | 2015 | 26345 | Bullock, AL | both |
| 388 | 20126 | 01013 | 2015 | 33475 | Butler, AL | both |
| 438 | 115285 | 01015 | 2015 | 33522 | Calhoun, AL | both |

In [22]: `combo[combo.TimePeriod == "2015"].tail(10)`

Out[22]:

| | Population | GeoFips | TimePeriod | IncomePC | GeoName | _merge |
|--------|------------|---------|------------|----------|----------------|--------|
| 153057 | 8334 | 56043 | 2015 | 42573 | Washakie, WY | both |
| 153105 | 7230 | 56045 | 2015 | 44841 | Weston, WY | both |
| 153153 | 14710229 | 91000 | 2015 | 60088 | New England | both |
| 153201 | 49083944 | 92000 | 2015 | 56284 | Mideast | both |
| 153249 | 46742511 | 93000 | 2015 | 45571 | Great Lakes | both |
| 153297 | 21095876 | 94000 | 2015 | 47123 | Plains | both |
| 153345 | 82092717 | 95000 | 2015 | 42649 | Southeast | both |
| 153393 | 40234946 | 96000 | 2015 | 44876 | Southwest | both |
| 153441 | 11710907 | 97000 | 2015 | 46320 | Rocky Mountain | both |
| 153489 | 55225488 | 98000 | 2015 | 53014 | Far West | both |

And notice that the data is giving me county level stuff, but also aggregated values.

Summary

It looks like I have the data to answer my questions. Lots more work to do, but progress!