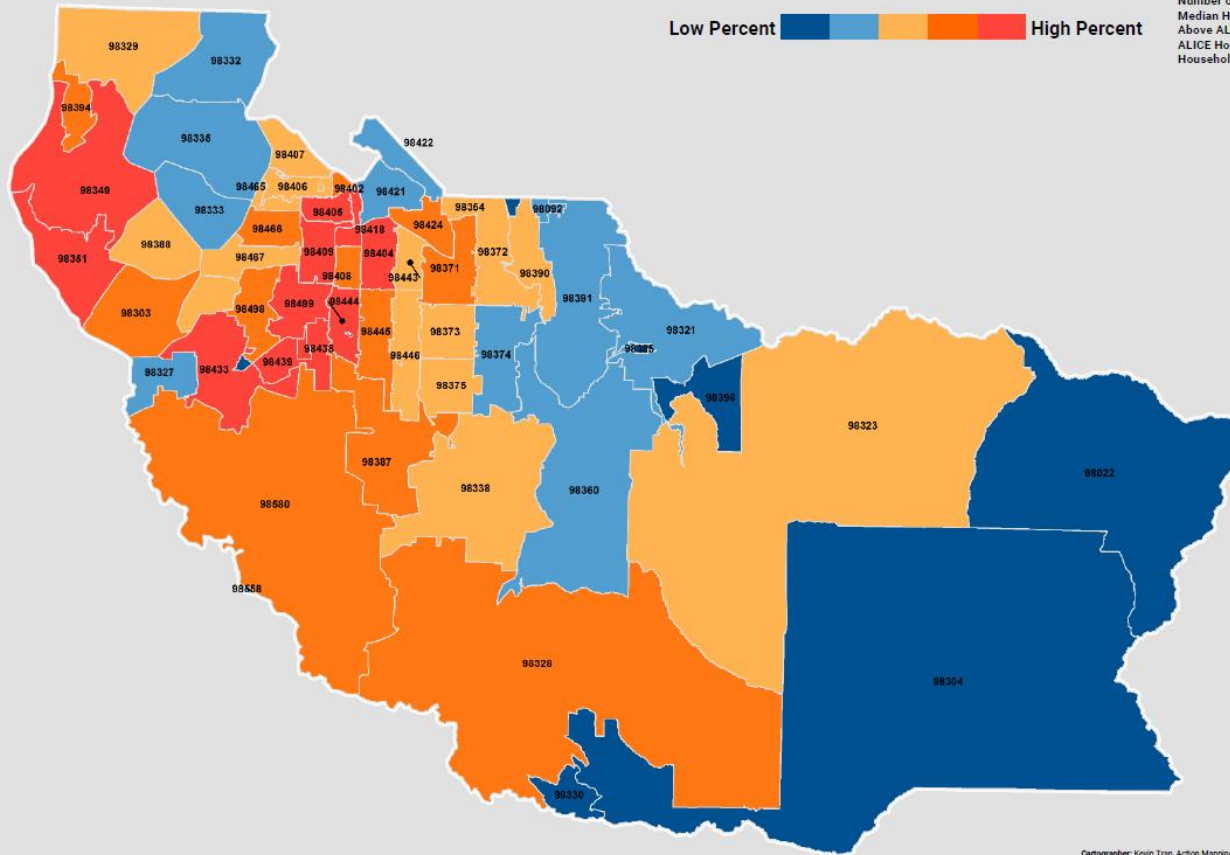


MANIPULATING CENSUS INCOME DATA WITH PYTHON AND ARCGIS API

Kevin Tran

PERCENTAGE OF ALICE FAMILIES IN PIERCE COUNTY (2016)

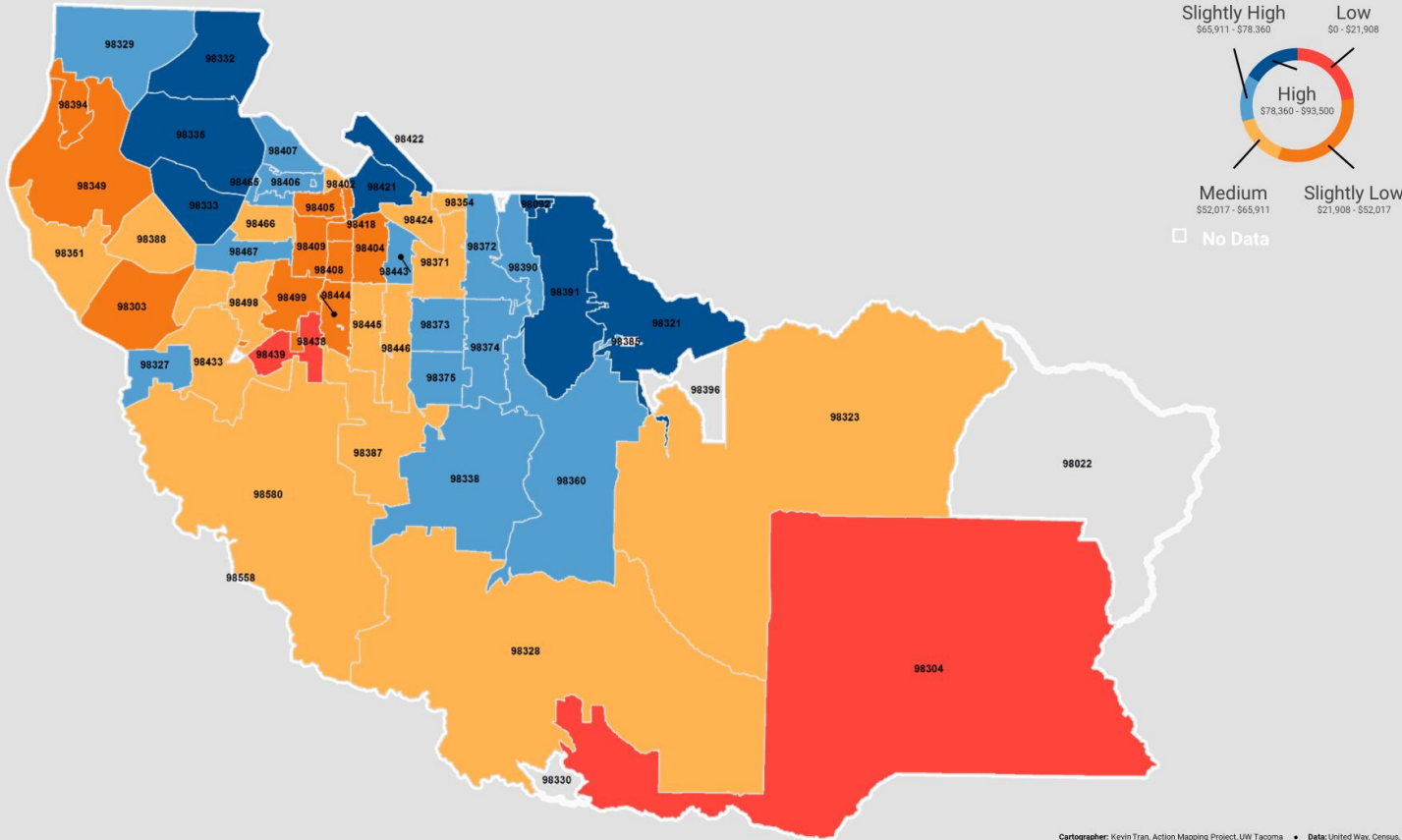


Background

As a project administrator for Action Mapping Project (A.M.P.), I was tasked with designing maps for United Way. My capstone is inspired by the work I did.

Question: Is there a way to automate data manipulation and visualization of Census data?

MEDIAN FAMILY INCOMES IN PIERCE COUNTY (2016)



Cartographer: Kevin Tran, Action Mapping Project, UW Tacoma • Data: United Way, Census, Pierce County

Background

Python libraries:

- Pandas
- Geopandas
- Zipfile
- ArcGIS API

Materials:

Census Data with:

- Columns reflecting income brackets
- It must be a CSV
- It must have a column head that matches the column head of a shapefile and similar values if able from each.

TIGER/LINE Shapefiles

How does it work?

Step 1: Unzip your files.

```
# STEP 1

import pandas
from zipfile import ZipFile
# specifying the zip file name
file_name = "block_group_data.zip"

# opening the zip file in READ mode
with ZipFile(file_name, 'r') as zip:
    # printing all the contents of the zip file
    zip.printdir()

    # extracting all the files
    print('Extracting all the files now...')
    zip.extractall()
    print('Done!')

file_name = "block_groups.zip"

# opening the zip file in READ mode
with ZipFile(file_name, 'r') as zip:
    # printing all the contents of the zip file
    zip.printdir()

    # extracting all the files
    print('Extracting all the files now...')
    zip.extractall()
    print('Done!')
```

Step 2: Manipulate Data via Pandas.

```
# STEP 2

x = pandas.read_csv('ACS_16_5YR_B19101_with_ann.csv', header=1)
x_drop = x.drop(x.columns[[4,6,8,10,12,14,16,18,20,22,24,26,28,30,32,34,36]], axis = 1)
x_drop['Poverty_Count'] = x_drop.iloc[:,4:8].sum(axis=1)
x_drop['ALICE_Count'] = x_drop.iloc[:,8:15].sum(axis=1)
x_drop['ALICE_Beyond'] = x_drop.iloc[:,15:20].sum(axis=1)

new_table = x_drop[['Id', 'Id2', 'Geography', 'Poverty_Count', 'ALICE_Count', 'ALICE_Beyond']]
new_table.rename(columns = {'Id2':'GEOID10'}, inplace = True)

database = pandas.DataFrame(new_table)
database.GEOID10.astype('int64')
```

How does it work?

Step 3: Manipulate Data via GeoPandas then zip up files .

```
# STEP 3

gdf = geopandas.read_file('tl_2010_53053_bg10.shp')

shapefile = pandas.DataFrame(gdf)

shapefile['GEOID10']=shapefile['GEOID10'].astype('int64')

newgeo = shapefile.merge(database, on='GEOID10')

ALICE = geopandas.GeoDataFrame(newgeo, geometry='geometry')

# https://geopandas.org/io.html
ALICE.to_file('tl_2010_53053_bg10.shp', driver='ESRI Shapefile')

# //ZIP FILE//

def main():

    file_name = 'new_blocks.zip'

    # opening the 'Zip' in writing mode
    with zipfile.ZipFile(file_name, 'a') as file:
        # append mode adds files to the 'Zip'
        # you have to create the files which you have to add to the 'Zip'
        file.write('tl_2010_53053_bg10.cpg')
        file.write('tl_2010_53053_bg10.dbf')
        file.write('tl_2010_53053_bg10.shp')
        file.write('tl_2010_53053_bg10.shp.xml')
        file.write('tl_2010_53053_bg10.shx')
        print('Files added to the Zip')

    # opening the 'Zip' in reading mode to check
    with zipfile.ZipFile(file_name, 'r') as file:
        print(file.namelist())

if __name__ == '__main__': main()
```

Step 4: Publish shapefiles using ArcGIS Python API

```
from arcgis.gis import arcgis

gis = arcgis.GIS("https://www.arcgis.com", username="ENTER_USERNAME", password="ENTER_PASSWORD")

data = "new_blocks.zip"
shpfile = gis.content.add({}, data)
published_service = shpfile.publish()
display(published_service)
```


Outcomes:

Block Group Feature Layer

The screenshot shows the ArcGIS Online interface for an item named 'new_blocks'. The top navigation bar includes links for Home, Gallery, Map, Scene, Groups, Content, and Organization. The item page has tabs for Overview, Data, Visualization, Usage, and Settings. The Overview tab is active, displaying a thumbnail map of a region with a red outline. Below the thumbnail is a description of the data, which is a shapefile of census block groups. The description includes details about the data source (U.S. Census Bureau's Master Address File / Topologically Integrated Geographic Encoding and Referencing (MAF/TIGER) Database) and the data's structure (shapefile). The item is created by 'gentles_UWT_GIS' and was last updated on March 7, 2020. The item has a view count of 2. On the right side of the page, there are buttons for 'Open in Map Viewer', 'Open in Scene Viewer', 'Open in ArcGIS Desktop', 'Publish', 'Create View Layer', 'Export Data', 'Update Data', 'Share', and 'Metadata'. Below these buttons is a section for 'Item Information' with a progress bar and a 'Learn more' link. At the bottom, there is a 'Details' section with information about the item's source, creation date, and size.

Home Gallery Map Scene Groups Content Organization

new_blocks

Overview Data Visualization Usage Settings

Thumbnail

In order for others to use the information in the Census MAF/TIGER database in a geographic information system (GIS) or for other geographic applications, the Census Bureau releases to the public extracts of the database in the form of TIGER/Line Shapefiles.

Feature Layer (hosted) by gentles_UWT_GIS

Created: Mar 7, 2020 Updated: Mar 7, 2020 View Count: 2

Add to Favorites

Description

The TIGER/Line Files are shapefiles and related database files (.dbf) that are an extract of selected geographic and cartographic information from the U.S. Census Bureau's Master Address File / Topologically Integrated Geographic Encoding and Referencing (MAF/TIGER) Database (MTDB). The MTDB represents a seamless national file with no overlaps or gaps between parts; however, each TIGER/Line File is designed to stand alone as an independent data set, or they can be combined to cover the entire nation. Block Groups (BGs) are defined before tabulation block delineation and numbering, but are clusters of blocks within the same census tract that have the same first digit of their 4-digit census block number from the same decennial census. For example, Census 2000 tabulation blocks 3001, 3002, 3003... 3999 within Census 2000 tract 12110.02 are also within BG 3 within that census tract. Census 2000 BGs generally contained between 600 and 3,000 people, with an optimum size of 1,500 people. Most BGs were delineated by local participants in the Census Bureau's Participant Statistical Areas Program (PSAP). The Census Bureau delineated BGs only where the PSAP participant declined to delineate BGs or where the Census Bureau could not identify any local PSAP participant. A BG usually covers a contiguous area. Each census tract contains at least one BG, and BGs are uniquely numbered within census tract. Within the standard census geographic hierarchy, BGs never cross county or census tract boundaries, but may cross the boundaries of other geographic entities like county subdivisions, places, urban areas, voting districts, congressional districts, and American Indian / Alaska Native / Native Hawaiian areas. BGs have a valid code range of 0 through 9. BGs coded 0 were intended to only include water area, no land area, and they are generally in territorial seas, coastal water, and Great Lakes water areas. For Census 2000, rather than extending a census tract boundary into the Great Lakes or out to the U.S. nautical three-mile limit, the Census Bureau delineated some census tract boundaries along the shoreline or just offshore. The Census Bureau assigned a default census tract number of 0 and BG of 0 to these offshore, water-only areas not included in regularly numbered census tract areas.

Layers

Open in Map Viewer

Open in Scene Viewer

Open in ArcGIS Desktop

Publish

Create View Layer

Export Data

Update Data

Share

Metadata

Item Information

Low High

Top Improvement: Add terms of use

Details

Source: Feature Service

Created from: new_blocks, Shapefile

Data Last Updated: Mar 7, 2020, 4:21:06 PM

Size: 2 MB

Shared with: The item is not shared.

The screenshot shows the ArcGIS Online map viewer interface. The top navigation bar includes links for Home, new_blocks, and Kevin. The map viewer has tabs for Details, Add, Basemap, and Analysis. The 'new_blocks' layer is selected and visible on the map. The map shows a region with a red outline, similar to the one in the first screenshot. Below the map, there is a table of data for the 'new_blocks' layer. The table has columns for various attributes, including 'FUNCSTAT10', 'ALAND10', 'AWATER10', 'INTPTLAT10', 'INTPTLON10', 'ID', 'Geography', 'Poverty_00', 'ALICE_Count', and 'ALICE_Rate'. The table contains two rows of data, representing different census tracts.

Home new_blocks Kevin

Details Add Basemap Analysis

Save Share Print Directions Measure Bookmarks Find address or place

Contents

new_blocks

new_blocks

Imagery with Labels

new_blocks (Features: 560, Selected: 0)

FUNCSTAT10	ALAND10	AWATER10	INTPTLAT10	INTPTLON10	ID	Geography	Poverty_00	ALICE_Count	ALICE_Rate
5	3,155,498.00	0.00	+47.3832347	-122.5944235	1500000053053072352	Block Group 2, Census Tract 725.05, Pierce County, Washington	0.00	100.00	198.00
5	26,973,708.00	3,698,514.00	+47.1289902	-122.6037942	15000000530530729031	Block Group 1, Census Tract 729.03, Pierce County	21.00	347.00	162.00

Full Screen

Full Screen

Full Screen

Outcomes:

County-Lv Feature Layer

Home Gallery Map Scene Groups Content Organization

new_counties

Overview Data Visualization Usage Settings

Open in Map Viewer

Open in Scene Viewer

Open in ArcGIS Desktop

Publish

Create View Layer

Export Data

Update Data

Share

Metadata

Item Information

Learn more

Low High

Top Improvement: Add terms of Use

Details

Source: Feature Service

Created from: new_counties, Shapefile

Data Last Updated: Mar 7, 2020, 6:18:02 PM

Size: 2 MB

Shared with: The item is not shared.

Description

The TIGER/Line Files are shapefiles and related database files (.dbf) that are an extract of selected geographic and cartographic information from the U.S. Census Bureau's Master Address File / Topologically Integrated Geographic Encoding and Referencing (MAF/TIGER) Database (MTDB). The MTDB represents a seamless national file with no overlaps or gaps between parts; however, each TIGER/Line File is designed to stand alone as an independent data set, or they can be combined to cover the entire nation. The primary legal divisions of most States are termed counties. In Louisiana, these divisions are known as parishes. In Alaska, which has no counties, the equivalent entities are the organized boroughs, city and boroughs, and municipalities, and for the unorganized area, census areas. The latter are delineated cooperatively for statistical purposes by the State of Alaska and the Census Bureau. In four States (Maryland, Missouri, Nevada, and Virginia), there are one or more incorporated places that are independent of any county organization and thus constitute primary divisions of their States. These incorporated places are known as independent cities and are treated as equivalent entities for purposes of data presentation. The District of Columbia and Guam have no primary divisions, and each area is considered an equivalent entity for purposes of data presentation. The Census Bureau treats the following entities as equivalents of counties for purposes of data presentation: Municipios in Puerto Rico, Districts and Islands in American Samoa, Municipalities in the Commonwealth of the Northern Mariana Islands, and Islands in the U.S. Virgin Islands. The entire area of the United States, Puerto Rico, and the Island Areas is covered by counties or equivalent entities. The 2010 Census boundaries for counties and equivalent entities are as of January 1, 2010, primarily as reported through the Census Bureau's Boundary and Annexation Survey (BAS).

Layers

new_counties

Open in Export To Print Settings Enable Attachments Service URL Metadata

Home new_blocks

Overview Data Visualization Usage Settings

Open in Map Viewer

Open in Scene Viewer

Open in ArcGIS Desktop

Publish

Create View Layer

Export Data

Update Data

Share

Metadata

Item Information

Learn more

Low High

Top Improvement: Add terms of Use

Details

Source: Feature Service

Created from: new_blocks, Shapefile

Data Last Updated: Mar 7, 2020, 6:18:02 PM

Size: 2 MB

Shared with: The item is not shared.

Description

The TIGER/Line Files are shapefiles and related database files (.dbf) that are an extract of selected geographic and cartographic information from the U.S. Census Bureau's Master Address File / Topologically Integrated Geographic Encoding and Referencing (MAF/TIGER) Database (MTDB). The MTDB represents a seamless national file with no overlaps or gaps between parts; however, each TIGER/Line File is designed to stand alone as an independent data set, or they can be combined to cover the entire nation. The primary legal divisions of most States are termed counties. In Louisiana, these divisions are known as parishes. In Alaska, which has no counties, the equivalent entities are the organized boroughs, city and boroughs, and municipalities, and for the unorganized area, census areas. The latter are delineated cooperatively for statistical purposes by the State of Alaska and the Census Bureau. In four States (Maryland, Missouri, Nevada, and Virginia), there are one or more incorporated places that are independent of any county organization and thus constitute primary divisions of their States. These incorporated places are known as independent cities and are treated as equivalent entities for purposes of data presentation. The District of Columbia and Guam have no primary divisions, and each area is considered an equivalent entity for purposes of data presentation. The Census Bureau treats the following entities as equivalents of counties for purposes of data presentation: Municipios in Puerto Rico, Districts and Islands in American Samoa, Municipalities in the Commonwealth of the Northern Mariana Islands, and Islands in the U.S. Virgin Islands. The entire area of the United States, Puerto Rico, and the Island Areas is covered by counties or equivalent entities. The 2010 Census boundaries for counties and equivalent entities are as of January 1, 2010, primarily as reported through the Census Bureau's Boundary and Annexation Survey (BAS).

Layers

new_blocks

Open in Export To Print Settings Enable Attachments Service URL Metadata

INTPTLONR	SX	Poverty_C0	ALICE_Count	ALICE_Revs	HL_Y	Geograph_1	Poverty_1	ALICE_Col_1	ALICE_Re_1
-120.7401446	0500000US3077	11,079.00	29,062.00	17,310.00	0500000US3077	Yakima County, Washington	11,079.00	29,062.00	17,310.00
-120.6185423	0500000US3007	2,537.00	8,197.00	7,485.00	0500000US3007	Chelan County, Washington	2,537.00	8,197.00	7,485.00
-122.6496358	0500000US3035	6,920.00	24,589.00	34,331.00	0500000US3035	Kitsap County, Washington	6,920.00	24,589.00	34,331.00

```

x_drop['ALICE_Count'] = x_drop.iloc[:,8:15].sum(axis=1)

x_drop['ALICE_Beyond'] = x_drop.iloc[:,15:20].sum(axis=1)

new_table = x_drop[['Id', 'Id2', 'Geography', 'Poverty_Count', 'ALICE_Count', 'ALICE_Beyond']]

new_table.rename(columns = {'Id2':'GEOID10'}, inplace = True)

database = pandas.DataFrame(new_table)

database.GEOID10.astype('int64')

database

```

	Id	GEOID10	Geography	Poverty_Count	ALICE_Count	ALICE_Beyond
0	8600000US83501	83501	ZCTA5 83501	1279	4537	3393
1	8600000US83822	83822	ZCTA5 83822	83	254	120
2	8600000US83824	83824	ZCTA5 83824	27	46	17
3	8600000US83832	83832	ZCTA5 83832	37	170	216
4	8600000US83851	83851	ZCTA5 83851	104	222	81
5	8600000US83855	83855	ZCTA5 83855	85	375	160
6	8600000US83870	83870	ZCTA5 83870	18	37	17
7	8600000US83872	83872	ZCTA5 83872	7	14	145
8	8600000US97031	97031	ZCTA5 97031	392	1975	1994



zip boundaries (Features: 598, Selected: 0)					
STATEFP10	ZCTA5CE10	GEOID10	CLASSFP10	MTFCC10	FUNCSTAT10
53	98822	5398822	B5	G6350	S
53	98821	5398821	B5	G6350	S
53	98357	5398357	B5	G6350	S
53	98663	5398663	B5	G6350	S
53	98684	5398684	B5	G6350	S
53	98848	5398848	B5	G6350	S
53	99166	5399166	B5	G6350	S
53	98851	5398851	B5	G6350	S
53	99103	5399103	B5	G6350	S

Strengths and Limitations

Strengths

- Simple and effective way to generate shapefiles and visualize data.
- Gets away with using ArcPy package.
- It is possible in theory when customizing code to manipulate and visualize other Census data. However, one must know what data and tools to mess with.

Limitations

- Not so much a limitation, but the code is roughly written. There is yet a way to store zip file names in a dictionary to use for re-zipping files.
- For this specific code, it presumes that column headers for shapefiles used do not change.

< One of the column headers need to change. This is manipulating zip boundary data.



THANKS FOR VIEWING!!!

Any questions?