

Introduction to GIS for Social and Environmental Analysis (CITY201)

Lab 6: Vector Operations, Part B

Purpose of the lab: To learn how to acquire data from the census website and to practice relational data model and vector analysis operations.

We will be using data from the **American Community Survey (ACS)**. The ACS is an ongoing survey by the US Census that provides vital information on a yearly basis about the United States and its people. The collected data helps address social and economic questions. In contrast to the decennial census that surveys all people in the United States, the ACS surveys a representative sample of the population. You can read more about the difference between the two surveys here: <https://www.census.gov/programs-surveys/acs/about/acs-and-census.html>

Next, click on the link below and browse through the ACS guide. Familiarize yourself with the ACS subjects, hierarchy of the major ACS geographic entities and ACS data releases (i.e. 1-years vs. 5-year estimates).

https://www.census.gov/content/dam/Census/programs-surveys/acs/about/ACS_Information_Guide.pdf

1. Download Tabular Census Data

Instead of using data from our 201_Data Folder, we will acquire data from sources available through the internet. We will first **acquire U.S Census tabular (attribute) data from the census website** then download shapefiles.

1. Go to <https://data.census.gov/cedsci/>
2. Click “Advanced Search,” then select ‘Geography’. For geographic type, look under “Most Commonly Used Geographies” and select ‘Block Group’. For state, select ‘Pennsylvania’. For county, select ‘Philadelphia’. Finally, select ‘All Block Groups within Philadelphia, County, Pennsylvania’ and click Search.

The screenshot displays the 'Advanced Search' interface on the Census Data website. It features three main filter panels: 'BROWSE FILTERS', 'WITHIN (STATE)', and 'WITHIN (COUNTY)'. The 'BROWSE FILTERS' panel on the left has 'Geography' selected. The 'WITHIN (STATE)' panel shows 'Pennsylvania' selected. The 'WITHIN (COUNTY)' panel shows 'Philadelphia County, Pennsylvania' selected. The 'WITHIN (TRACT)' panel shows 'All Block Groups within Philadelphia County, Pennsylvania' selected. The search results on the right list various census tracts within Philadelphia County, Pennsylvania.

BROWSE FILTERS	WITHIN (STATE)	WITHIN (COUNTY)	WITHIN (TRACT)
Topics	Alabama	Northumberland County, Pennsylvania	Within Other Geographies
Geography	New Hampshire	Perry County, Pennsylvania	☒ All Block Groups within Philadelphia County, Pennsylvania
Years	New Jersey	Philadelphia County, Pennsylvania	Census Tract 1, Philadelphia County, Pennsylvania
Surveys	New Mexico	Pike County, Pennsylvania	Census Tract 10, Philadelphia County, Pennsylvania
Codes	New York	Potter County, Pennsylvania	Census Tract 10.01, Philadelphia County, Pennsylvania
	North Carolina	Schuylkill County, Pennsylvania	Census Tract 10.02, Philadelphia County, Pennsylvania
	North Dakota	Snyder County, Pennsylvania	Census Tract 100, Philadelphia County, Pennsylvania
	Ohio	Somerset County, Pennsylvania	Census Tract 101, Philadelphia County, Pennsylvania
	Oklahoma	Sullivan County, Pennsylvania	Census Tract 102, Philadelphia County, Pennsylvania
	Oregon	Susquehanna County, Pennsylvania	
	Pennsylvania	Tioga County, Pennsylvania	
	Puerto Rico	Union County, Pennsylvania	
	Rhode Island		

3. A list of tables will pop up. If you don't see the Table you want, you can enter a topic name (like "population" or "median income") in the search bar. Select "Total Population" – **make sure you are getting the one from the ACS and NOT the Decennial Census. Also check that you are getting the right table number: "Total Population (B01003)," the right year (2019) under "Select Table Vintages," and the right Geography (All Block Groups within Philadelphia, County, Pennsylvania).**

TOTAL POPULATION

Survey/Program: American Community Survey

Years: 2019,2018,2017,2016,2015,2014,2013

Table: B01003

4. Click "Download Table"
5. Next place your cursor in the search bar at the top of the page – you should see a pop-up that shows your geography (block groups in Philadelphia" is still selected. Search for "median household income" and click on the table name "Median Household Income in the Past 12 months (in 2019 inflation-adjusted dollars)' (B19013)".
6. Click "Download Table"
7. **Save the zipped file to a location where you can access it from Excel. Note that Apporto does not have Excel installed.** You will have to work either on your own computer or on a computer in the lab. Then, extract the files to the location you can access from Excel.
8. Review the different files – see how they have data on both the variable "estimate" and the variable "margin of error"?

Margin of Error or MOE is a measure of the possible variation of an estimate around the population value. It is supposed to measure the maximum amount by which the sample results are expected to differ from those of the actual population. The margin of error is the range of values below and above the sample statistic in a confidence interval and is used to interpret the precision of the estimate. The Census Bureau publishes MOE's at a 90% confidence level. **So we have 90% confidence that the true population value falls within the range given by the combination of the sample estimate and the margin of error.**

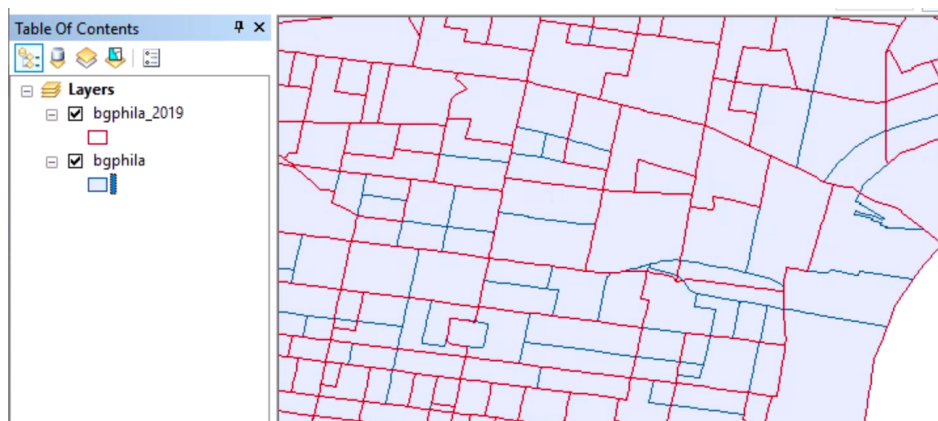
Look at the margin of error for some of the values in this data table. Take the margin of error (+/-), add it to and subtract it from the estimate to get the range. Do you see how big some of these ranges are? Now do you understand why I don't want you to use lots of decimal places when reporting this data?

2. Download Census Cartographic Boundary Files

Now that we have acquired the Census tabular block group data, we will **download the block group boundary files**.

1. Go to <https://www.census.gov/geographies/mapping-files.html>
2. On the timeline in the middle of the page, click on “2019”
3. Scroll to the bottom of the page and click on “TIGER/Line Shapefiles”
4. On the TIGER/Line page, click on “2019” on the timeline, then scroll to the bottom of the page and click “Web Interface”
5. Select “2019” and “Block groups” then click “Submit”
6. Now select “Pennsylvania” and click “Download”
7. Save the zipped file to your workspace for Lab 6. Then, extract the files to your workspace. I recommend keeping your spatial and tabular data in separate folders.
8. Repeat these steps to get 2019 census tracts instead of block groups.
9. You should end up with 2 folders:
 - a. tl_2019_42_bg
 - b. tl_2019_42_tract

Make sure you are getting 2019 versions for both files! There are significant differences in the 2019 and 2020 block group boundary files:



Note that there is an option to download all boundary files for a geographic area at once, but it makes a very large download file (for Pennsylvania, it's 2.23 GB, even when zipped).

3. Prepare Your Tabular Data

Before starting to edit your tabular files, make a copy of all of them and put them into a separate folder as a backup (call the folder 'original_data' or some such).

Open the file “ACSDT5Y2019.B01003-Data” in Excel.

1. Rename the column headers so they are meaningful to you (e.g. totpop). Make sure no special characters (, - &) or spaces are in the column headings (underscore_ is ok).
2. Delete the entire second row (**NOT the column, the row!**).
3. Make sure data values start in row 2.
4. We need to clean up the GEO_ID column so that it has the right ID number to be able to join the data to the block group shapefile.
 - a. See how the first part of the text string is the same on every row – “1500000US”? That part of the GEO_ID does not match the way the Geo IDs are stored in the shapefile. So we are going to remove that part.
 - b. Add a new column to the right of the GEO_ID column and label it GEOIDTR (the TR stands for TRIM).
 - c. In the first data row of the GEOIDTR column (cell B2), type this formula: `=RIGHT(A2, 12)`
 - d. Hit enter. See how it gives you just the numbers after the string “1500000US”? This formula returns the right most characters from the string in the cell in the formula for the numbers entered in the formula. Try changing the number “12” to some other number to see what you get. When you have finished exploring, change it back to 12.
 - e. Now we need to fill this formula down through the whole set of the data. I find it easiest to scroll to the bottom of the set of data (Shortcut: Ctrl + Shift + Down Arrow), select the last cell I want to choose, then scroll back up to the cell that has your formula (Shortcut: Ctrl + Shift + Up Arrow). Now you can use the “Fill Down” command to fill the formula down the whole column. Put your cursor in any cell to see how the references in the formula have changed.
 - f. Convert calculations into values: Here’s another great Excel trick. Select the whole column with our calculated values and then select copy. Now go to Edit>Paste Special, and select Values. Click ok. Looks the same, right? But now put your cursor in one of the cells where you had calculated the values. Wow! No more calculations – just numbers.
5. Delete the columns we don’t need: GEO_ID (the one with the long string that starts “1500000US”) and NAME
6. Click on any one cell within the table.
7. Go to Save As and for Save as type choose ‘CSV (Comma delimited) (*.csv)’.
8. Save the new file in your workspace using the file name ‘phbg_totpop.csv.’

Repeat the steps for the other table that contains median household income data for block groups in Philadelphia (Median Household Income in the Past 12 Months in 2019 Inflation-Adjusted Dollars) (“ACSDT5Y2019.B19013_data_with_overlays”) and name your file “phbg_medhhi.csv.”

Note that median household income estimates are missing for a number of block groups. That is because the Census suppresses data if reliable estimates of a variable are not available for a particular geographic unit or other data quality or confidentiality concerns warrant restriction of the data. This occurs most often at the block group level, the smallest unit of aggregation for ACS data.

Close Excel.

4. Prepare Your Spatial Data

4a. Select geography for just Philadelphia

Open ArcGIS and add the shapefiles you downloaded from the Census website to the map. Add the two tables you created: phbg_totpop.csv and phbg_medhhi.csv

Review the attribute table of the shapefile as well as the two other tables. Note that the shapefile is for all of Pennsylvania, but our table data is just for Philadelphia County. Next we want to make a new shapefile of block groups for just Philadelphia.

Look in the shapefile Attribute Table. Note that the GEOID field is a unique ID number for each block group. This number is a concatenated number that combines 4 other fields: STATEFP, COUNTYFP, TRACTCE, BLKGRPCE. The state code for Pennsylvania is 42. **What's the code for Philadelphia County?** [Hint: use your identify tool to find the info in one of the block groups in Philadelphia.]

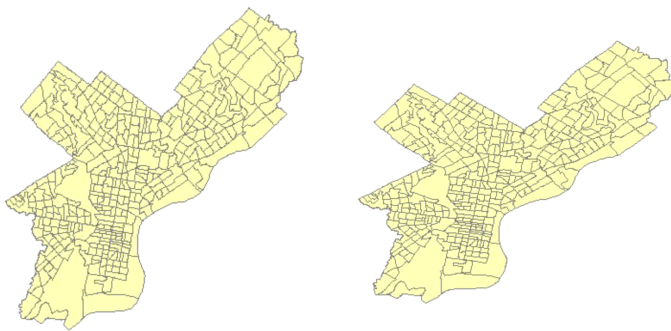
Use Select by Attribute to select all of the block groups in Philadelphia based on the County code, then export that selection to a new shapefile (I called mine bgphila.shp).

Repeat this process to create a shapefile of all tracts in Philadelphia (I called mine trtphila.shp).

4b. Get projections into UTM

Use the ArcGIS toolbox to change the projection for both the block group shapefile and the tract shapefile to NAD_1983_UTM_Zone_18N.

Use the preview function in ArcCatalog to check that your projections look right:



UTM

NOT UTM

EITHER open a new ArcGIS file and add both UTM files and both tables to the map OR keep working in the same ArcGIS file but check the properties of your Map to make sure the map view is set to the proper projection (ArcGIS sets the view projection for a map based on the first layer that is added to the map).

4c. Set up your shapefiles for joining

Open the table for total population and check the Properties of your GEOIDTR field – what is its data type?

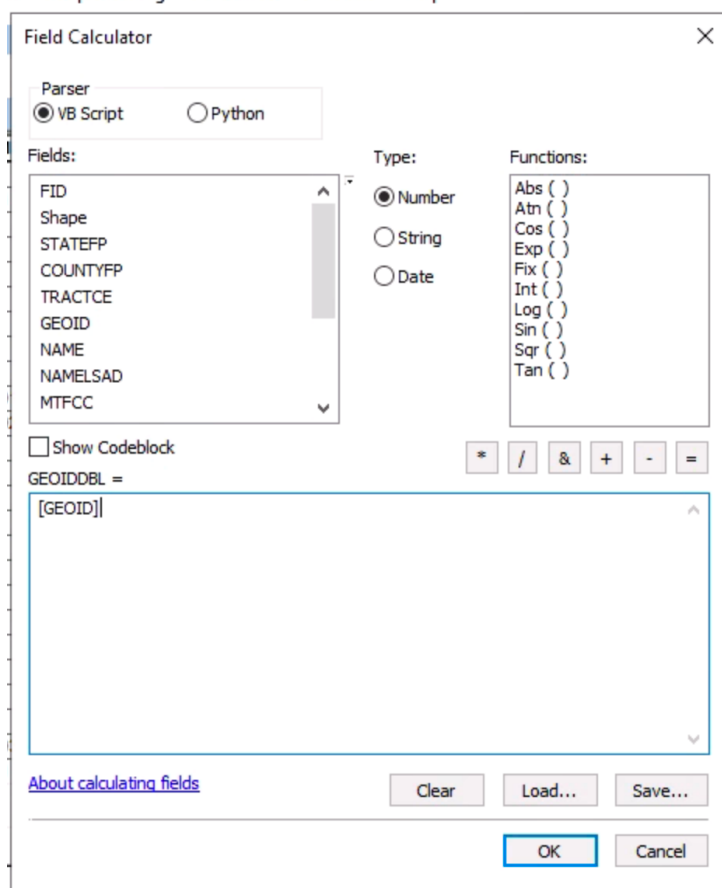
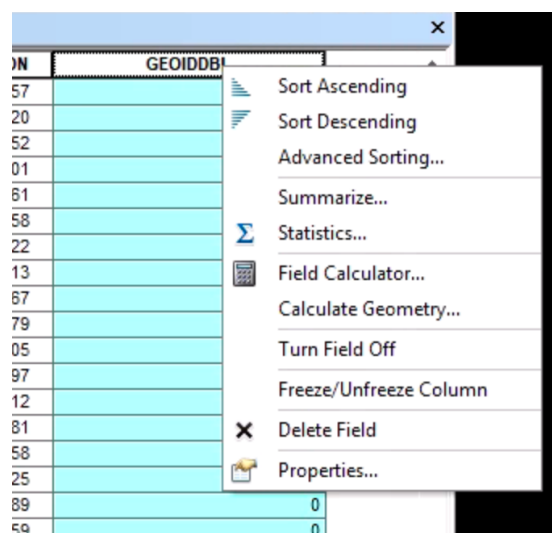
Now open the attribute table for one of your shapefiles and check the properties of the GEOID field – what is its data type?

These two fields – GEOID in the shapefile and GEOIDTR in the table – look like they have the exact same data in them, but they are defined as different field types. Do you think they will join? Try it.

When ArcMap brings in a .csv table, it sets the Field properties based on what kind of data is in the field. For your total population table, it saw a list of numbers and defined the data type as “Double,” which is one of ArcMap’s number formats. The census shapefiles have the GEOID stored as a “String,” which is a text format. These two fields won’t join.

There are probably other ways to solve this problem, too, but my workaround is to **add a new field to the shapefile and store the GEOID in a Double format:**

1. Open the shapefile attribute table.
2. Add a field – name it GEOIDDBL. Set the data type to “Double.”
3. Use the Calculate Field to calculate the GEOIDDBL Field – make it equal to the GEOID field and make sure you have set the type to Number.
4. Click OK



Now you have a field in your shapefile that your table data will join to. Repeat these steps for the other shapefile (do this for both the block groups and tracts shapefiles).

5. Join the Spatial and Attribute Data

Now **join the two tables (one at a time) onto the shapefile** using the fields 'GEOIDTR' and 'GEOIDDBL'.

Review the attribute table of your block group shapefile. It should now include total population and median household income fields.

6. View Block Group and Tract Data

Make a choropleth map of median household income by block group.

Move the tracts layer above the block groups layer in the table of contents so that you can see the tracts boundaries on top of the block groups. Set the color of the tracts layer to hollow and increase the thickness of the outline width.

Do you see a lot of variation in median household income within individual tracts?

7. Explore Margin of Error by Calculating Standard Errors for Median Household Income

Now let's explore the impact of the Margin of Error on the variability of median household income data. Remember: Margin of Error or MOE is a measure of the possible variation of an estimate around the population value. The margin of error (MOE) and a known confidence level is used to interpret the precision of the estimate. The Census Bureau publishes MOE's at a 90% confidence level. This implies a 10 percent chance of incorrect inference for all estimates. Also remember that the Census suppresses data if reliable estimates of a variable are not available for a particular geographic unit or other data quality or confidentiality concerns warrant restriction of the data, so median household income estimates are missing for a number of block groups.

7a. Select geography for just the records that have median household income data

First, go into your block group shapefile and **select the records that have median household income data**. You should have 1172 block groups. **Make a new feature layer or shapefile with these selected block groups.**

7b. Calculate the Standard Error for Median Household Income

Now we will calculate the Standard Error for Median Household Income. A standard error (SE) measures the variability of an estimate due to sampling. It provides a quantitative measure of the extent to which an estimate derived from a sample can be expected to deviate from the value for the full population.

In your shapefile that has only the block groups with median household income data, **create a new field called "MHHI_SE" and set the data type to float. Calculate the Standard Error for Median Household income in this field** (Remember the tool to calculate a new field?) by entering the formula $MEDHHI_ME / 1.645$ (that is, the margin of error for median household income divided by 1.645).

7c. Explore variation in the Standard Error for Median Household Income

Use a chloropleth map to explore the variation in this value – how much variation is there in the reliability of the estimates for different block groups?

Now do you see why we shouldn't use super precise decimal points with this kind of data?

8. Download and Prepare Tract Data

For the assignment, you will also download population data at the census tract level.

1. Go to <https://data.census.gov/cedsci/>
2. Click "Advanced Search," then select 'Geography'. For geographic type, look under "Most Commonly Used Geographies" and select 'TRACT'. For state, select 'Pennsylvania'. For county, select 'Philadelphia'. Finally, select 'All Tracts within Philadelphia, County, Pennsylvania' and click Search.
3. A list of tables will pop up. Select "Total Population" – **make sure you are getting the one from the ACS** and NOT the Decennial Census. Also check that you are getting the right table number: "Total Population' (B01003)". Also check that you are getting the right year (2019) under "Select Table Vintages."
4. What's this? Because there are fewer tracts than block groups, the census website is now showing you a preview of the table. Does this table look like it is set up the same way as the other tables? Does it look like it will be easy to join to a tract shapefile? NO.
5. Click "Customize Table"
6. In the middle of the toolbar, click "Transpose"
7. Hmm. This still doesn't look right, but let's try it.
8. Click "Download Table"
9. Save the zipped file to your workspace. Then, extract the files to your workspace. Move them to the folder with your other table data.
10. Review the file – whew! This looks right.
11. Clean the data with the same steps you used before:
 - a. Rename the column headers so they are meaningful to you (e.g. totpop). Make sure no special characters (., - &) or spaces are in the column headings (underscore_ is ok).
 - b. Delete the entire second row (NOT the column, the row!).**
 - c. Make sure data values start in row 2.
 - d. Clean up the GEO_ID column so that it has the right ID number to be able to join the data to the block group shapefile.
 - e. Add a new column to the right of the GEO_ID column and label it GEOIDTR (the TR stands for TRIM).

- f. In the first data row of the GEOID_TR column (cell B2), paste this formula: `=RIGHT(A2, 11)` – NOTE that we only need 11 digits this time instead of 12, because the GEOID does not include the Block Group number.
- g. Fill this formula down through the whole set of the data.
- h. Convert calculations into values: Select the whole column with our calculated values and then select copy. Now go to Edit>Paste Special, and select Values. Click ok.
- i. Delete the columns we don't need: GEO_ID (the one with the long string that starts "1400000US") and NAME
- j. Click on any one cell within the table.
- k. Go to Save As and for Save as type choose 'CSV (Comma delimited) (*.csv)'.
- l. Save the new file in your workspace using a file name that makes sense to you.

ASSIGNMENT

Lab #3-6: Must do 1 Full Lab Report & 3 Slim Lab Reports –IF YOU DID NOT DO A FULL LAB REPORT FOR LABS 3-5, THIS IS YOUR LAST CHANCE TO GET CREDIT FOR THIS REQUIREMENT

Objective

The objective of this assignment is to identify an area in which to build a new 'luxury' food market in the city of Philadelphia. The market must be:

1. within a Census block group having a median household income value in the top 10% of all block groups containing population.
2. within a Census tract with a population density in the top 10% for all tracts containing population.
3. in a contiguous region with an area of at least 110,000 square meters

Deliverables

Turn in a lab report addressing this objective. In your Methods section, clearly list the steps you used for this analysis. The report should include a map of the potential area(s) in which to build the market. An additional map or maps should be used to illuminate your methodology (showing in different layers the areas that meet the 3 criteria). *Don't forget that you should indicate in some way tracts that have no population and/or have no median income data.*

Lab 6 Summary of Operations

There were no new GIS operations as a part of Lab 6. Rather, you practiced previous GIS operations and learned the following tools:

- Census: Find and download table data
- Census: Find and download cartographic boundary files
- Census: Prepare table data for joining to spatial data

Therefore, there is no Summary of Operations table for this Assignment.