

## Joining Flat Files with Vector Layers

Looking at (one aspect of) the digital divide in Pennsylvania.

1. Get these data sets, and save them into your designated folder for QGIS files; extract both folders:
  - o PA Census Tracts from [TIGER/Line Shapefiles](#) - 2018, Web Interface, Census Tracts, Pennsylvania
  - o [American FactFinder](#) - Internet Access (under Advanced Search, Topics -> Housing -> Physical Characteristic), limit Geographies to Census Tracts in Pennsylvania (all) (B28002)
2. Start a new project (and then save it in your designated folder for QGIS files).
3. Add the Census Tracts shapefile as a layer.
4. Have a look at its attribute table.
5. Clean up the Internet Access (ACS\_17\_574\_B28002\_with\_ann) file in Excel, to remove data we don't want. (Open it, Save As something meaningful, and then make changes.)
  - o It's vital that we keep the GEO.id2 field, but let's change it to GEO\_id2.
  - o Let's say we're looking at the estimated number of households without internet access, column HD01\_VD13 (second to last column); we should keep that. We'll also want the estimated total number of households (HD01\_VD01). We should keep those two columns, plus GEO\_id2. We should also give them more useful names: total\_households and no\_internet\_households.
  - o We should also remove that explanatory row (row 2).
6. Now, we can add it as a delimited text layer. Unlike the tree data file, this one has no geometry. After it's added, change its name to something easier to work with, like "census\_data"
7. To add the information from this layer to the attribute table of our census tracts, we need to perform what's called a "Join." Right-click on the census tracts layer, and choose "Properties," and then choose "Joins" on the left-hand menu.
8. Hit the plus sign.
9. Since we're in the Properties of the tracts layer, the join layer will be the census\_data layer. We're combining these tables based on the census tract ID, so GEO\_id2 will be our join field, and GEOID will be our target field.
10. Click OK, and go look at the attribute table now. (Aren't you glad we changed the layer name on our census data? That said, we could have changed the column label prefix in the layer join dialog box, too, but it's useful to have it match.)
11. Now we're going to add a column to hold the percentage of households without internet. (I called mine "no\_net\_per.") It's done just like the density of trees in the previous exercise, only we take the number of households without internet and divide by the total number of households in each tract. (Don't forget to make it a decimal type column, not a whole number type.)
12. After we add that column, sort it. Nulls aren't great, right? I propose we highlight those rows and use the Field Calculator to set them to zero.
13. Now we're going to make a graduated color map, just like we did in the first example.
14. We're also going to set the display values on mouse-over.

What can you tell me about the digital divide in PA? How about in Allegheny County?