Hands-on Session 2: Combining summary data and microdata (QGIS)

Linking IPUMSI Data to Geography Shapefiles in QGIS

For QGIS version 3.30.3, released 05/26/2023 – (New versions come out periodically, to install the new version the previous version will need to be manually un-installed.)

1. Downloading QGIS:
   a. The QGIS main website is available at https://qgis.org/en/site/index.html
      i. Currently QGIS is available for download in 25 languages and provides accompanying documentation.
      ii. IPUMSI languages for QGIS – English, French, Spanish, Portuguese, Arabic, Russian, Chinese but not Swahili
   b. Click on the “Discover QGIS” tab on the top ribbon.
   c. Under “Discover” click the button that says “Get started”
   d. Here you can find the installation button, documentation downloads, and user forums for specific questions. Click on “Get the installer” to download the software.
      i. The next page will ask you to specify which operating system your computer uses, and after selecting the appropriate option, the installer should take you through the final steps.

2. Prepping files for GIS
   a. Table prep
      i. The table downloaded from SDA will need a few adjustments so that it can be read into GIS software. This simply involves removing unnecessary information and items that may confuse the program.
      ii. First, remove all metadata information from the top rows of the table.
      iii. Clear the contents from the first column of the table and move the second column into the space left in the first column.
      iv. Highlight the first column and using the “Text to Columns” feature, select “fixed width” and align the arrow which will make the split with the beginning of the text representing the labels. Click finish and this should separate the geography codes from their associated labels.
      v. Now that we have separate columns for codes and labels, assign them labels in the first row. We recommend using words like “GEOLEVEL1, geo1, or geocode” for the code column, and “ADMIN_NAME, label, or province” for the label column.
      vi. Lastly, make sure to remove any final rows that show column totals, they will not be necessary for mapping.

Final Table example:
b. Shapefile prep
   i. All shapefiles provided by IPUMS-International can be found at
      https://international.ipums.org/international/gis.shtml
   ii. For this exercise, we used the harmonized geography for Vietnam so the
       corresponding shapefile will be under “Spatially harmonized first-level/second-
       level geography”
   iii. Clicking on the link shows you a table of each country in IPUMS-International,
       the title of each administrative division, the corresponding geographic variable,
       the corresponding harmonized migration variable(s) if available, and the
       associated shapefile.
   iv. Clicking on the last link in the table downloads the shapefile in a .zip file format,
       which will then need to be unzipped by right clicking on the file and selecting
       “extract” (for Mac users, simply double click on the .zip file to extract the
       necessary files).
   v. 8 files will appear, as each of these files provide information needed to open,
       view, and edit shapefiles. All 8 files will have the same name, and only the file
       type will differ, and all files need to be in the same folder location to properly
       load the shapefile into any software package.
       Folder Location example:

---

<table>
<thead>
<tr>
<th>Name</th>
<th>Date modified</th>
<th>Type</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capture.png</td>
<td>5/22/2023 4:20 PM</td>
<td>PNG file</td>
<td>93 KB</td>
</tr>
<tr>
<td>Capture2.png</td>
<td>5/22/2023 4:42 PM</td>
<td>PNG file</td>
<td>13 KB</td>
</tr>
<tr>
<td>Diff_Mean.csv</td>
<td>5/22/2023 3:46 PM</td>
<td>Microsoft Excel CSV</td>
<td>2 KB</td>
</tr>
<tr>
<td>Diff_Percent.csv</td>
<td>5/22/2023 3:17 PM</td>
<td>Microsoft Excel CSV</td>
<td>2 KB</td>
</tr>
<tr>
<td>geo1_vn1989_2009.CPG</td>
<td>5/22/2023 3:00 PM</td>
<td>CPG file</td>
<td>1 KB</td>
</tr>
<tr>
<td>geo1_vn1989_2009.dbf</td>
<td>5/22/2023 3:59 PM</td>
<td>DBF file</td>
<td>13 KB</td>
</tr>
<tr>
<td>geo1_vn1989_2009.prj</td>
<td>5/22/2023 3:00 PM</td>
<td>PRJ file</td>
<td>1 KB</td>
</tr>
<tr>
<td>geo1_vn1989_2009.shp</td>
<td>5/22/2023 3:00 PM</td>
<td>SHP file</td>
<td>3,763 KB</td>
</tr>
<tr>
<td>geo1_vn1989_2009.shx</td>
<td>5/22/2023 3:00 PM</td>
<td>SHX file</td>
<td>17 KB</td>
</tr>
<tr>
<td>geo1_vn1989_2009.zip</td>
<td>5/22/2023 3:00 PM</td>
<td>Compressed (zipped)</td>
<td>3,812 KB</td>
</tr>
<tr>
<td>QGIS_cohort_sheet.docx</td>
<td>5/22/2023 4:41 PM</td>
<td>Microsoft Word Document</td>
<td>110 KB</td>
</tr>
</tbody>
</table>

---

3. Opening files in QGIS
Hands-on Session 2: Combining summary data and microdata (QGIS)

a. Opening the program will show several windows with relevant information. The pane furthest to the left shows the file browser where shapefiles and tables can be found. The middle pane will show any projects that have been saved or edited recently. Lastly, the pane furthest to the right will show saved templates that can be used to create a new project.

b. Tip: QGIS recommends creating an individual project for each map project or map series, but opening a blank project will still open a new space to add in shapefiles and tables.

c. Tip: After opening a project, the base map screen will be available, with the map space in the large area to the right of the screen, the folder browser tab on the left-hand side, and the layer tab on the bottom of the left-hand side.

   i. The folder browser is defaulted to show all available file locations on the computer to show both local files and any accessible remote servers. To access the local files on your computer click on the folders titled C:\, Users, and finally your individual username.

   ii. The layers tab functions like a table of contents, showing all current objects added to the map. Files can be dragged into the map space or layers tab to be added to the map, users can drag objects up and down in the layers tab to change the drawing order, and users can right-click on the objects to remove them.

d. After navigating to the folder containing the shapefiles and tables in the folders tab, the easiest way to import the shapefiles is to click and drag the file into the map space or the layer tab.

e. This displays the boundaries and areas identified in the IPUMS International shapefiles, and right clicking on the file name in the layers tab and clicking on “Open Attribute Table” will open the table showing data attributed to each geographic unit. This is the data used to join data to the shapefile.
Hands-on Session 2: Combining summary data and microdata (QGIS)

f. Adding the table requires a few more steps, as the default settings in QGIS are not suitable for dragging and dropping the table into the map space. To add the table, click on the “Open Data Source” icon in the upper left corner above the file browser window. After opening the window, navigate on the left hand panel to “Add delimited text”.

```
<table>
<thead>
<tr>
<th>File name</th>
<th>C:sERS\home003\Downloads\NDIRA\CH_Means.csv</th>
</tr>
</thead>
<tbody>
<tr>
<td>Layer name</td>
<td>CH_Means</td>
</tr>
<tr>
<td>Encoding</td>
<td>UTF-8</td>
</tr>
</tbody>
</table>
```

i. The file name is the entire file path to the table that is to be joined. The button on the far right showing three dots will open a separate window that can allow users to navigate to the file in the folder system.

ii. Under “Geometry Definition”, select no geometry. This will only need to be done once, as QGIS will remember settings for future maps. All other settings can remain the same, and the only features that need to be changed are the field types which are the drop down menus under each of the field names.

iii. The first column is the key that will link the table to the shapefile, so it is **IMPORTANT** that the field type of this field matches the field type of the GEOLEVEL1 field in the shapefile. IPUMS International distributes all shapefile fields in Text format, so changing the key field in the table to Text is the easiest joining method.

iv. Lastly, users should make sure that any data fields, in this case the “value” field, is set to integer fields if dealing with whole numbers, or decimal fields if dealing with decimal numbers.

v. Clicking “Add” at the bottom of the window will then add the table to the layer tab on the bottom left on the screen.

vi. **FOR MAC USERS**: To access the Add Delimited Text window, navigate to the menu bar on the top of the screen to select “Layer”, choose “data source manager”, and finally select delimited text on the left hand side.
Hands-on Session 2: Combining summary data and microdata (QGIS)

- Now that both items are in the map and formatted correctly, they can be joined by right clicking on the shapefile, clicking on “Properties”, and navigating to the “Joins” panel on the left-hand side. Click the green plus arrow on the bottom of the window to add a new join, and specify the table to be joined, the field of the table to be used as a joining key, and the field of the shapefile that matches the key, then hitting “Ok”.
- To check if the join completed correctly, open the attribute table as before, and see if the fields of the table are now included in the subsequent attribute table.

4. Changing the Symbology
   - To change the appearance of the shapefile to showcase the data that has been joined, right click on the shapefile in the layers tab, click on “Properties”, and navigate to the “Symbology” tab on the left-hand side.
   - This is the page to change the colors or patterns of individual geographic units, boundaries, or to assign graduated symbology to show variation of a specific value.
   - To change the symbology to reflect the data fields, first change the symbology method at the top of the window from “Single Symbol” to “Graduated”.
   - Then select the data field in the “Value” line underneath.
   - Select a color scheme in the “Color Ramp” line in order to show a range of colors from one end of the data range to the other.
   - Finally, at the bottom of the window, select a method to classify the data under “Mode”. If nothing appears in the large white space under “classes”, hit the “Classify” button under where you selected the classification mode. This will add any classes that are missing.
     - Classification breaks can be edited manually to user specified breaks, but we recommend using the classification methods first to observe the range of the data being mapped. To then change the classes, double-click on the class to be edited, and simply type in the breaks desired.
     - The default classification system in QGIS is the “Quantile” system, which places the data into classes which contain equal numbers of data points.
Hands-on Session 2: Combining summary data and microdata (QGIS)

g. Click on “Ok” at the bottom of the window to apply the new symbology, and the data should now be displayed in a choropleth map.