LAB 6 – CALCULATING & QUERYING ATTRIBUTE DATA: PART 2

What you'll Learn: This is the second part of a two-part lab where we learn how to query, calculate, and rank attribute data. You will learn how to apply an attribute join and perform simple queries in and beyond the field calculator.

Data: Data produced in Lab 5.

What You'll Submit: 1) Answers to all questions in a .doc or .pdf, and 2) a final draft map (in .pdf or .png) of Kentucky Distressed Block Groups that shows ranked areas of economic distress based on the latest unemployment rate, educational attainment rate, and per capita market income (PCI), with consideration of urban vs. rural areas.

Background: In the previous lab, we concluded by ranking distress based on a combination of key socioeconomic indicators, and we produced a draft of a map representing those rankings. Some issues arose, however. Urban areas (Lex, Louisville, Covington) were very tough to discern. As such, we couldn't easily figure out patterns of distress in Kentucky's biggest cities. This lab will walk through the final stages of analyzing our attribute data, and finish with making a couple of maps that help visualize urban areas.

Lab naming conventions: Tools that you click will be bolded, e.g., **QGIS Menu > File > New** to create a new QGIS project file. Text that you'll type will have quotes around it, such as "MyNewProject.qgs" and names of existing datasets and directories will be *italicized*, e.g., *DataToUse.zip*. Key terms will be underlined. **Important tips and key instructions will be in bold red font.**

STEP 1: CALCULATING RURAL AND URBAN BLOCK GROUPS

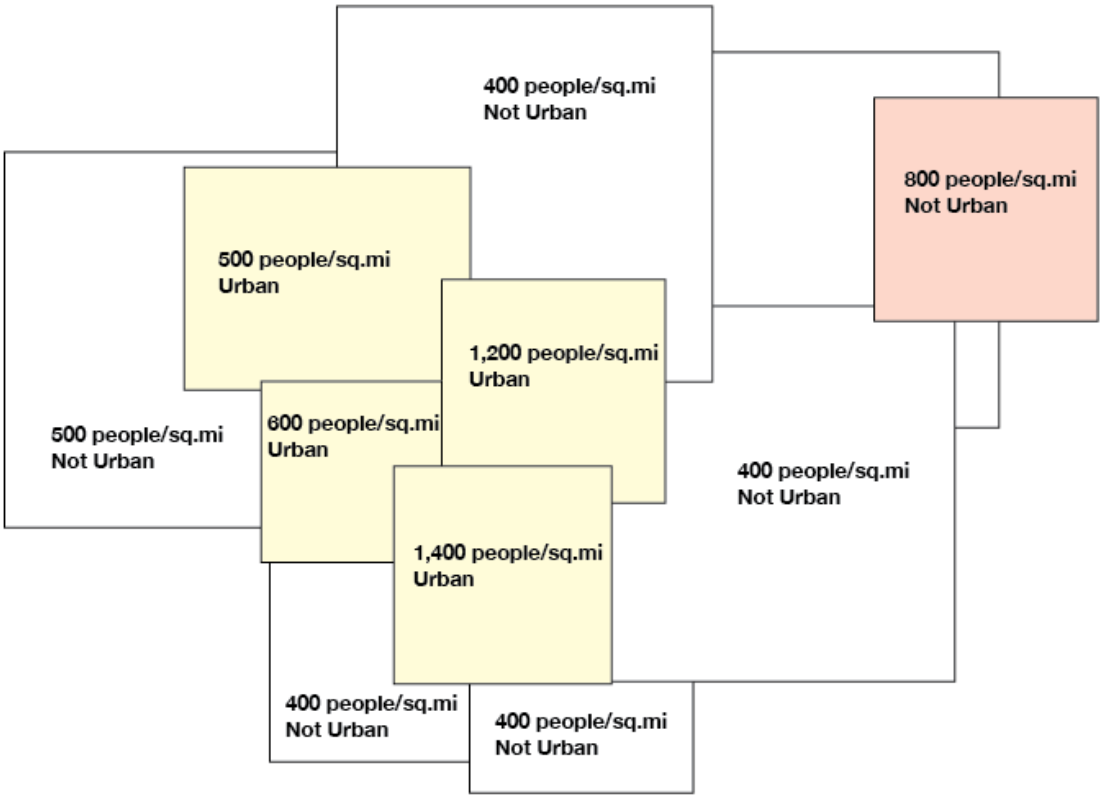
Begin by creating a new workspace. **I recommend the following steps:**

- Create a folder for "Lab6"
- Open your *Lab5.qgs* map document
- **Click** "Save as…" and save it as "Lab6" in the *Lab6* folder
- **This way, you have a new document with functional file paths and will be picking up exactly where you left off.**

Considering the issues we had with visualizing cities in the last lab, we want to figure out how to more easily distinguish between rural and urban block groups. The Census Bureau defines a block group is urban if its population density is 1,000 people per square mile or greater. Also, any neighboring block group that has 500 people per

square mile or greater is also an urban tract.

Consult the following illustration:



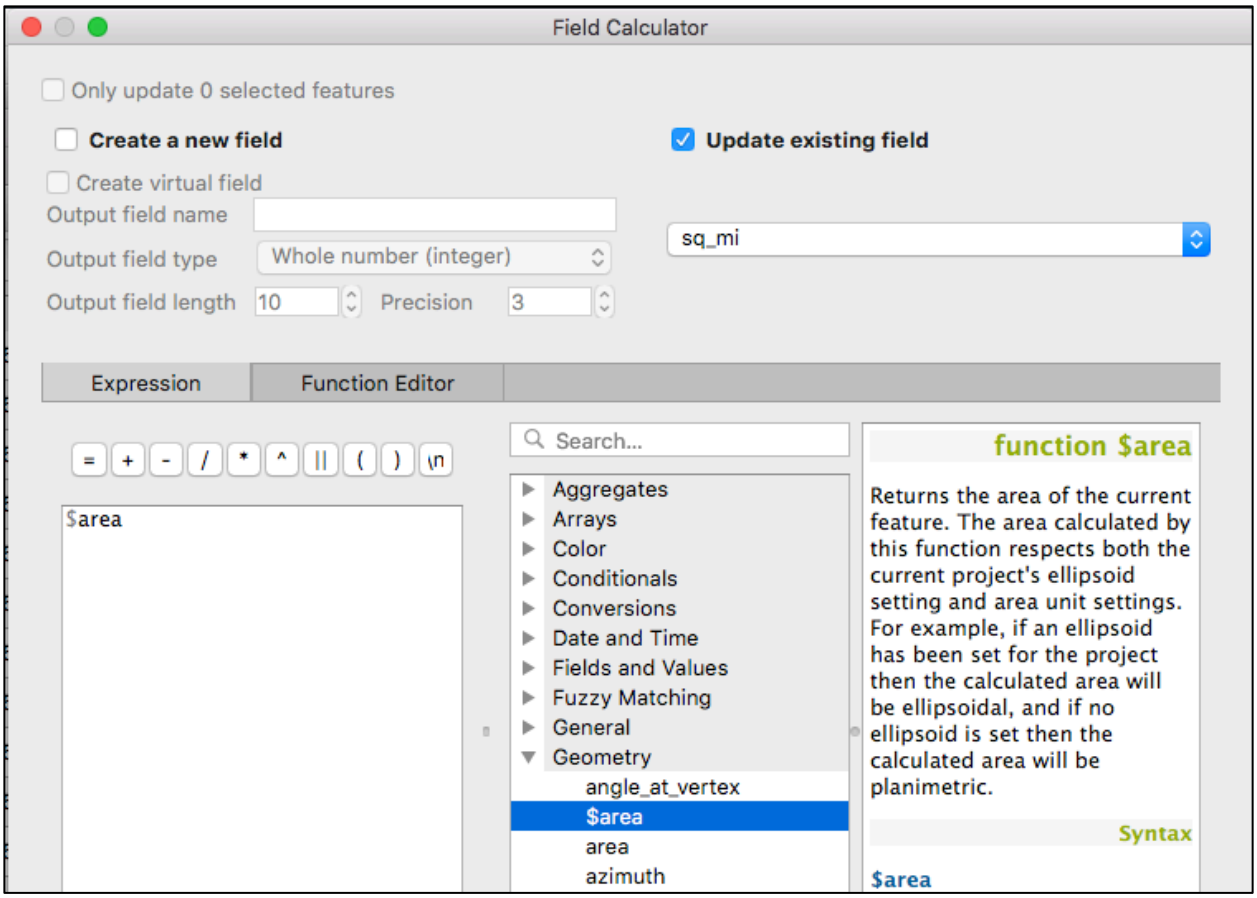
Open the attribute table for your layer *KyDistressBG* and create two new fields:

a. Field one should be a **Decimal** data type. Name it "Sq_Mi" and make the length "20" and the precision "5" – this field will calculate the area each tract.

b. Field two should be a **Whole Integer** data type. Name it "Urban" and make the length "1"– this field will record whether a block group is urban (1) or rural (0).

In this section we'll be calculating population density per block group. First, though, we must calculate area in the "Sq_Mi" field. You may recall we did this in Lab 3 and it was a bit tedious. Thankfully, there is an easier way to do it.

Open the **Field Calculator** and check **Update existing field.** From the drop-down bar, select "Sq_Mi". In the **Expression dialog window**, enter "$area" as seen below:

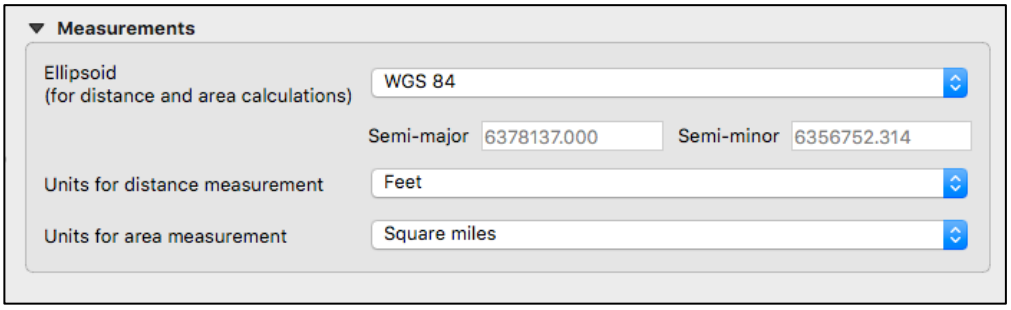


Press **OK** and you should see a set of values populate in the "Sq_Mi" field.

We want to calculate square miles, so the numbers should look something like the values to the right – but they probably don't. Why?

sq_mi
0.02569
0.03717
0.03814
0.05520
0.05543
0.05701
0.05737
0.05737
0.05838
0.05871
0.05926
0.06445

The "$area" function calculates each feature's area **in the current project's projection, including its standard units**. We can change this by navigating to **Project > Properties > General,** and adjusting the settings as such:



After you've adjusted the Project Property measurement settings, **be sure to run the "$area" function in your Field Calculator again**. This

will recalculate the proper values, telling QGIS to calculate the area in square miles as opposed to square meters, feet, or whatever other value may have been entered.

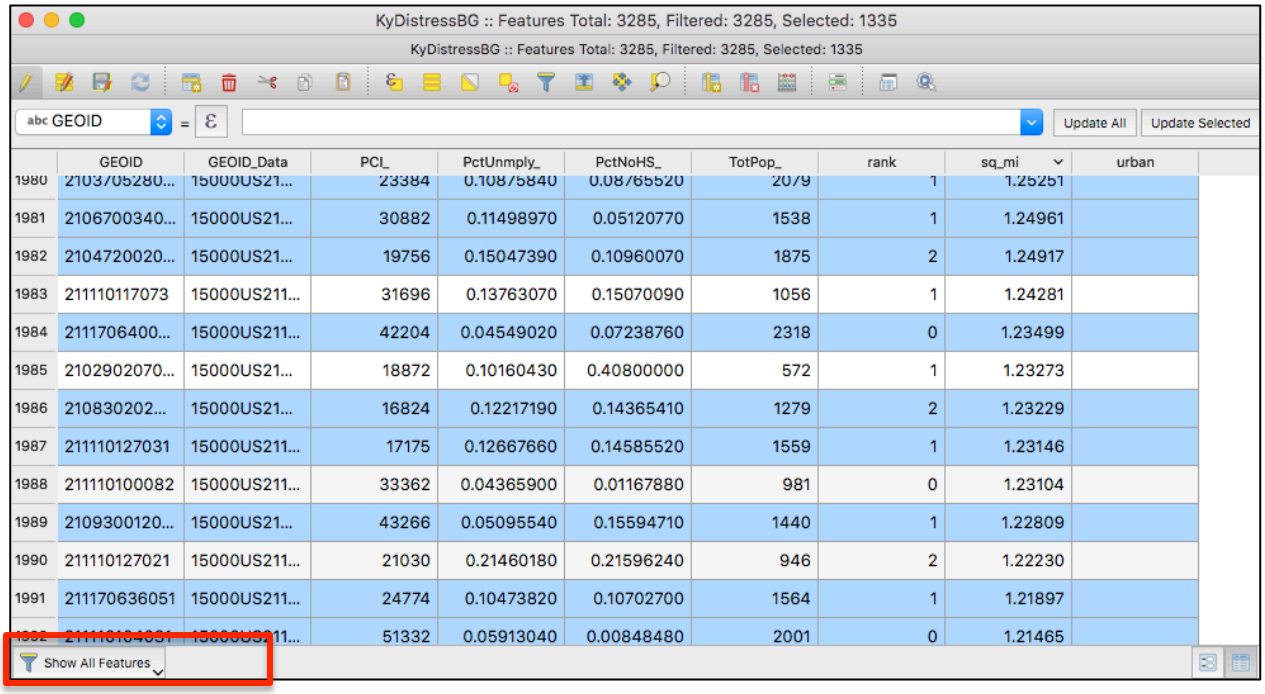
You might be thinking, "WOW… this was so much easier than running the **Add Geometry Tool** a bunch of times." But there was a reason Lab 3 asked you to run that tool over and over.

> **QUESTION 1:** Why did Lab 3 ask us to run the **Add Geometry** tool, whereas here, we can simply calculate the area in the field calculator?

Recall that in this step we are going to designate block groups as "urban" or "rural." Next, open the **Select By Expression** dialog and type the following:

"TotPop_" / "Sq_Mi" >= 1000

Execute the selection query. This should select 1335 records in the *KyDistressBG* attribute table:



	GEOID	GEOID_Data	PCI_	PctUnmply_	PctNoHS_	TotPop_	rank	sq_mi	urban
1980	2103705280...	15000US21...	23384	0.10875840	0.08765520	2079	1	1.25251	
1981	2106700340...	15000US21...	30882	0.11498970	0.05120770	1538	1	1.24961	
1982	2104720020...	15000US21...	19756	0.15047390	0.10960070	1875	2	1.24917	
1983	211110117073	15000US211...	31696	0.13763070	0.15070090	1056	1	1.24281	
1984	2111706400...	15000US211...	42204	0.04549020	0.07238760	2318	0	1.23499	
1985	2102902070...	15000US21...	18872	0.10160430	0.40800000	572	1	1.23273	
1986	210830202...	15000US21...	16824	0.12217190	0.14365410	1279	2	1.23229	
1987	211110127031	15000US211...	17175	0.12667660	0.14585520	1559	1	1.23146	
1988	211110100082	15000US211...	33362	0.04365900	0.01167880	981	0	1.23104	
1989	2109300120...	15000US21...	43266	0.05095540	0.15594710	1440	1	1.22809	
1990	211110127021	15000US211...	21030	0.21460180	0.21596240	946	2	1.22230	
1991	211170636051	15000US211...	24774	0.10473820	0.10702700	1564	1	1.21897	
			51332	0.05913040	0.00848480	2001	0	1.21465	

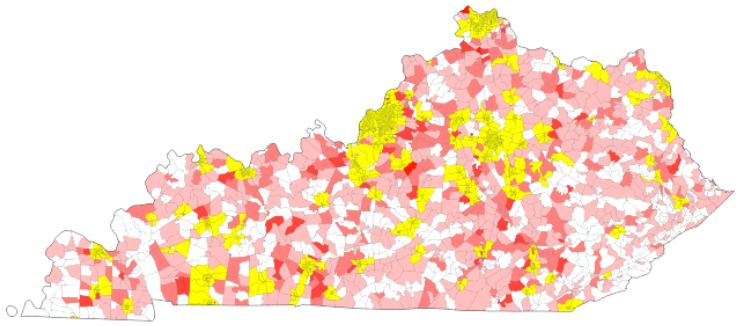
If this selects 0 records, pause. A couple of things might be going on. First, confirm that your values have been properly calculated. Did they resemble the "sq_mi" values in the screenshot above? If not, double-check your work. If the values *do* seem correct, it could be a problem with the Select by Expression query (QGIS can be finicky here). As an alternative, you can **Filter** the attribute table by clicking on the button in the red box above, choosing **Advanced Filter**, and entering the expression "TotPop_" / "Sq_Mi" >= 1000 there instead.

Once you have the appropriate 1335 records selected, **Calculate** them as Urban by inserting a “1” (without quotes) in your field calculator, and clicking **OK**. Be sure you have checked “Only update 1335 features” and “Update existing field” > “urban”.

Now that we have our Urban records calculated, navigate from the top toolbar **Vector > Research Tools > Select by location…** This tool allows us to select records based on a set of spatial parameters, as opposed to parameters within the data. In the **Select by location** window, match the following parameters:

- Select features from = *KyDistressBG*
- Where the features = “intersect”
- By comparing to the features from = *KyDistressBG*
- **Selected features only should be checked**
- Modify current selection by = creating new selection

Once your parameters are set, click **OK**. Your attribute table should now contain a selection of about 1886 records and resemble the following (note the colors might look different, as my data is currently a choropleth map of population):



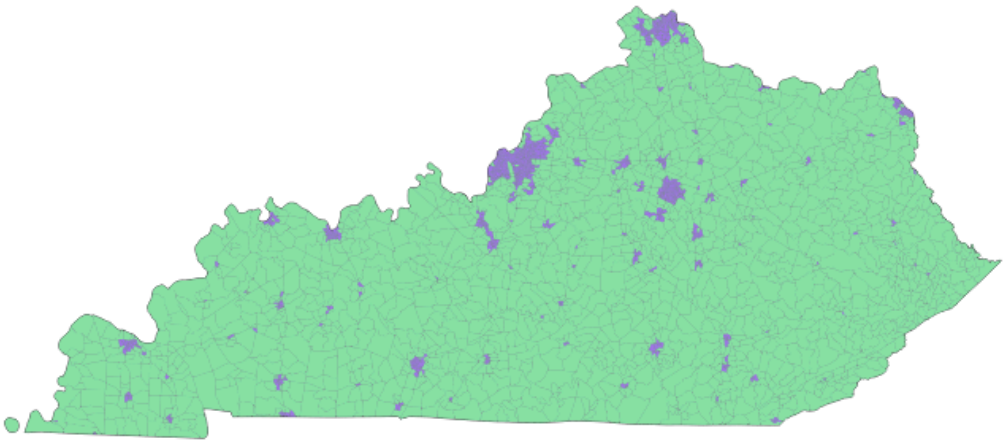
Now that we have this selection, we’ll complete another query to select only those where total population divided by square miles is greater than or equal to 500. (Recall that according to the diagram on page 2, any neighboring block group that has 500 people per square mile or greater is also an urban tract.) Navigate to **Select by Expression > Select from Current Selection** and enter:

“TotPop_”/”Sq_Mi” >= 500

Click **OK** and you should see the selected records jump to 1565. However, if you scroll through the attribute table, you’ll notice selections with Urban values of “NULL”. Open

the **Field Calculator** and calculate the selected records as “1” (without quotes).

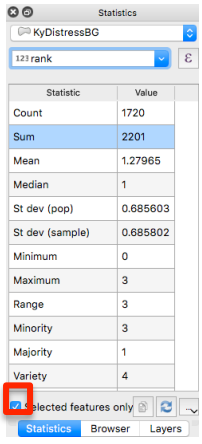
Finally, calculate the remaining records as “0” (without quotes) using whatever method you like (e.g., invert selection > Field calculator, resort the attribute table by “urban” field, etc). When you categorize the *KyDistressBG* layer by “urban”, it should now resemble the following:



QUESTION 2: Reclassify the *KyDistressBG* layer such that it is “Categorized” by the “urban” field. Paste a screenshot of that image into your Word document, making rural areas green and urban areas yellow.

QUESTION 3: In 2-3 sentences, describe what you have done using these selection queries. Specifically, why did we need to select using both the Select by Expression and the Select by Location tools?

STEP 2: IS DISTRESS MORE RURAL OR URBAN?



This section will follow the guiding question posed in the title: namely, in Kentucky, is distress a more rural or urban phenomenon?

We are mainly going to use the **Statistics** tool, located in the main toolbar at the Σ button, and also accessible via **View > Panels > Statistics.** The goal is to create a table that summarizes how many block groups are Rural or Urban and in which Distress Rank they fall. The goal is to make a chart that clearly shows our results. We use a custom Excel Spreadsheet to finish this task.

First, open the **Statistics** window and set the layer to *KyDistressBG* and set the field to "rank". Your summary table should resemble the figure to the right. Second, select all instances of "urban" = 0 in your attribute table. **Be sure that the "Selected features only" box is checked!**
Once you've selected all instances of "urban" = 0 in your attribute table, you should have 1720 records selected.

Next, we're going to perform a couple of selection queries based on all of these rural records. In the **Statistics** tool, click on the white space and change "rank" to "rank = 0", and press **Enter**. You should see the values change in the table: now, we are no longer looking at summary values for all 1720 records, but rather, summary values for only those records where "rank = 0".

Now, let's pause the QGIS work, and open a new spreadsheet in MS Excel (or equivalent program). We'll use Excel to copy out some important values and turn them into a handy chart. Begin by replicating the spreadsheet below:

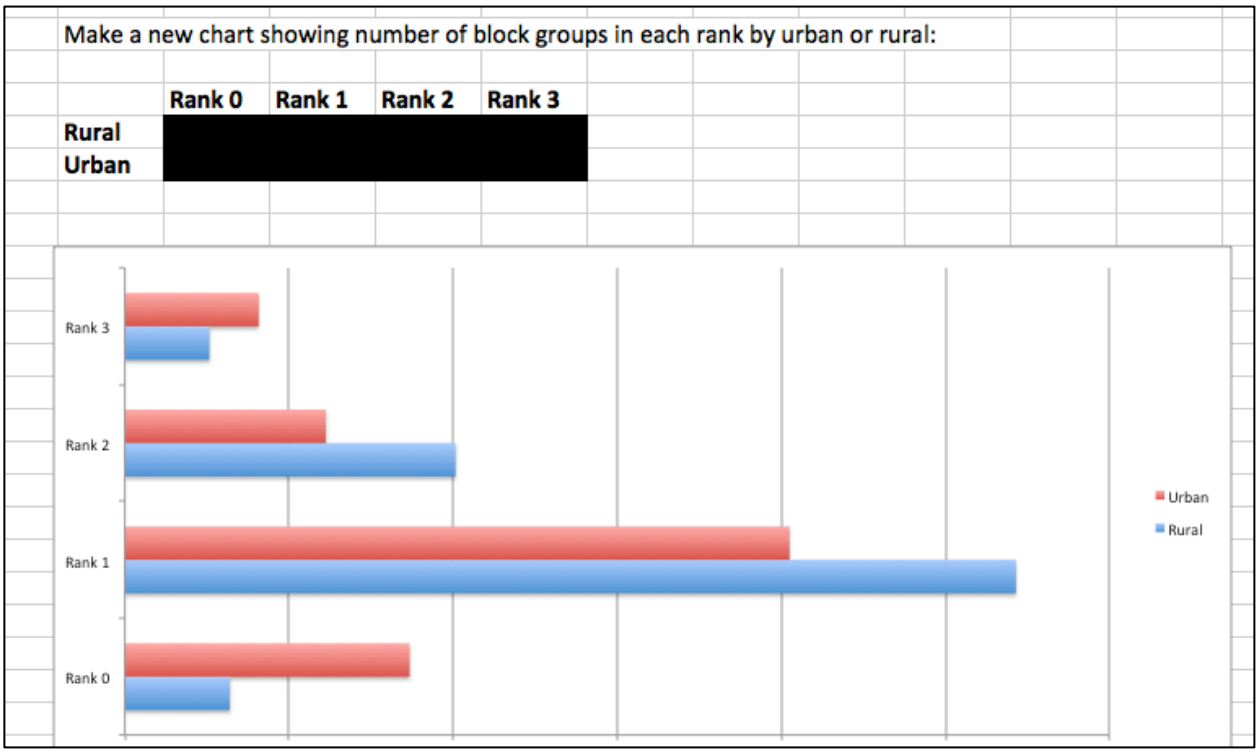
	Make a new chart showing number of block groups in each rank by urban or rural:				
		Rank 0	**Rank 1**	**Rank 2**	**Rank 3**
	Rural				
	Urban				

Once you've finished doing this – and I recommend saving the Excel spreadsheet in your workspace with an appropriate name (e.g., "rural-urban-rank") – we're going to return to the **Statistics** tool in QGIS. In the "Value" field for "Sum," you should see the value 128. This value refers to the total instances of records where the rank was equal to 0 in your selected records (e.g., rural records). So, you can type 128 into the cell where "Rural" aligns with "Rank 0", as such:

Make a new chart showing number of block groups in each rank by urban or rural:				
	Rank 0	**Rank 1**	**Rank 2**	**Rank 3**
Rural	128			
Urban				

Complete the rest of the seven values by updating your query in the **Statistics** tool. **Note that once you finish the values for Rank 1, 2, and 3, you'll have to select "Urban" values (where "urban" = 1) in the *KyDistressBG* Attribute Table!** You can double-check your work by adding up the values for "Rural" and "Urban" in Excel, and cross referencing them against the total number of rural (0) and urban (1) records in the *KyDistressBG* Attribute Table in QGIS.

Excel has tools built in that allow us to generate charts on the fly. Navigate to the **Charts** tab in Excel and select **Insert > Charts > Column**. This should automatically generate a decent looking chart of rural versus urban distress, similar to the one below:



QUESTION 4: Paste a screenshot of your Excel spreadsheet, mimicking my screenshot above. Note that I've excluded values from my chart. Your chart should look similarly to mine in shape. Also note that I am running a very old version of Excel on my laptop and yours may look different. It's okay if your chart is oriented differently, as long as the values are right!

QUESTION 5: In 2-3 sentences, what are some conclusions that you can draw about the urban vs. rural distribution of distress in Kentucky? What do you think might be some explanations for your conclusions?

STEP 3: MAPPING DISTRESSED AREAS IN KENTUCKY

Now it's time present a map of our work. The map will show the whole state, with insets for Louisville, Covington, and Lexington, all at the same scale. We will use "Distress" or "Rank" field to symbolize our areas.

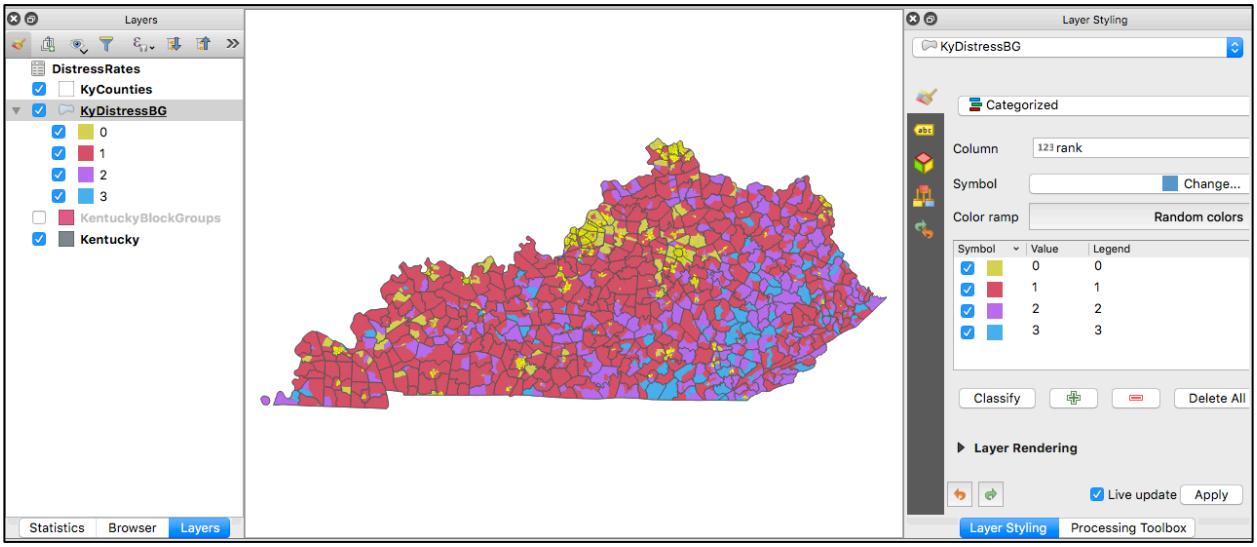
STEP 3.1: SELECTING A COLOR SCHEME

Let's start. Your QGIS project should include at least 4 layers:

- Kentucky
- Kentucky block groups
- Kentucky distress block groups
- Kentucky counties

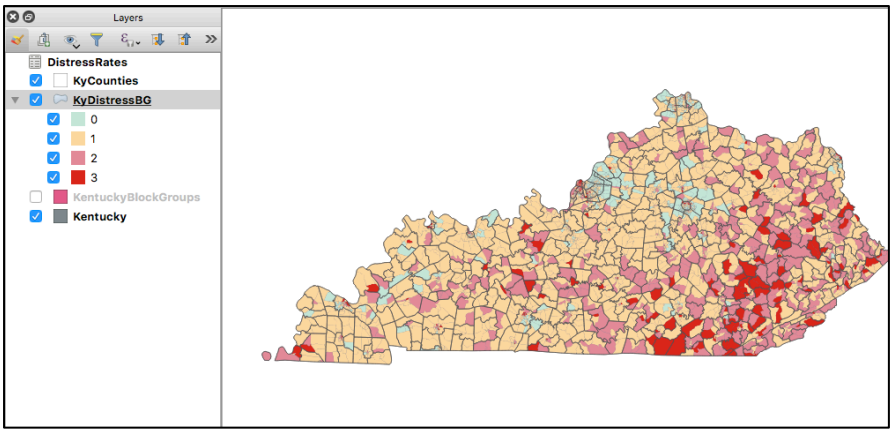
Click on the layer *KyDistressBG* and navigate to the **Layer styling panel** or **Properties > Symbology**. Select "Categorized" from the menu, then set the Column to "rank" and click the **Classify** button.

You should see the following:



Note that you can feel free to deselect or remove any spurious classes that pop up in the Classification window by unchecking it, or clicking the red **Minus** button.

Now, let's make a symbol color scheme and custom legend. While the color choices are up to you, there are some basic principles to observe for this map.



Since Rank 0 indicates an economic level better than the national average, it should have

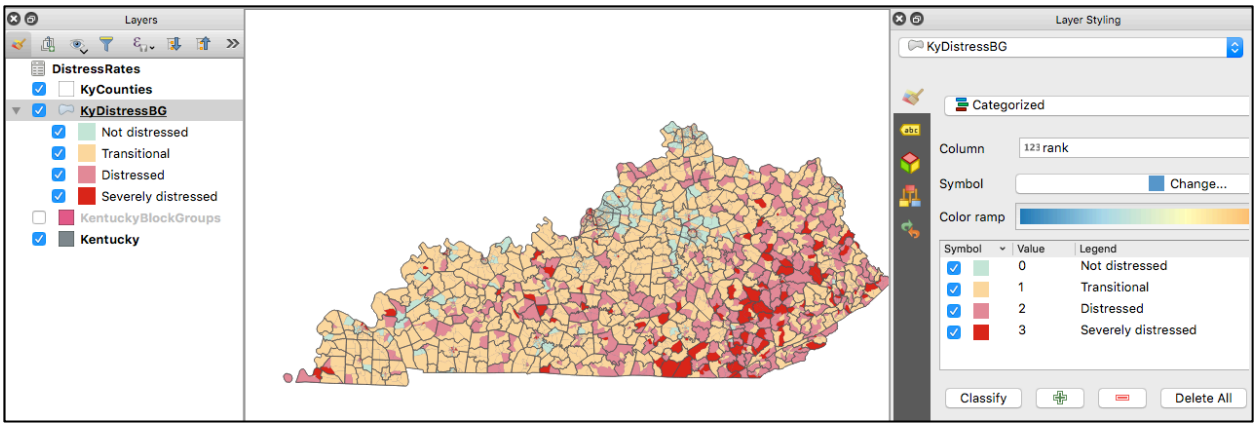
a different color than the other Ranks. Ranks 1-3 should indicate magnitude, since they represent increasingly distressed economic conditions.

Using what you know of map design and symbolization in QGIS, you may either 1) develop your own color scheme that follows the instructions above, or 2) mimic to the best of your ability the color scheme below:

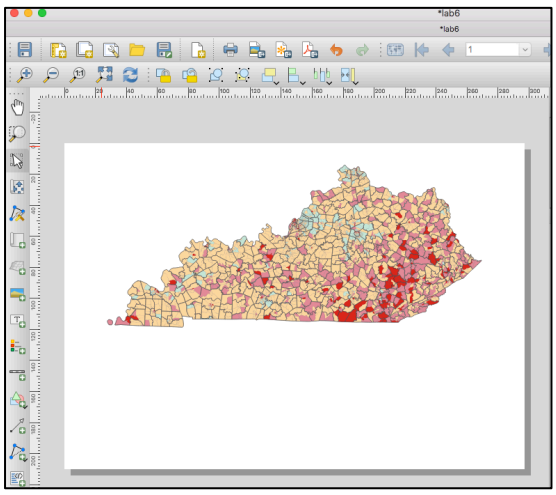
Next, in the **Symbology** tab or the **Layer styling panel**, give each class a recognizable name in the legend. You can do this by simply double-clicking on the legend and entering in your new title. Change the classes to the following:

- **Rank 0** should become “Not Distressed”
- **Rank 1** should become “Transitional”
- **Rank 2** should become “Distressed”
- **Rank 3** should become “Severely Distressed”

If working in the **Layer styling panel**, the final legend should resemble this:



STEP 3.2: FINAL MAP LAYOUT AND INSET MAPS



The remainder of this lab will take place in the **Print Composer**. Create a new print layout by selecting **Project > New Print Layout…** and title it “Lab6”.

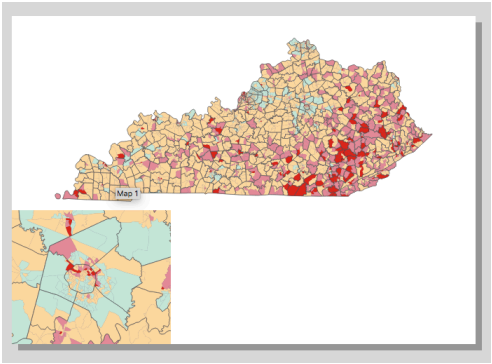
Begin by **right+clicking** the *KyDistressBG* layer and selecting **Zoom to layer**. This will ensure that when we add our data to the print layout, it is properly scaled.

Navigate to your print layout screen and add the data by dragging a box that covers roughly the upper two-thirds of the empty canvas. Adjust the scale as needed (this can be done

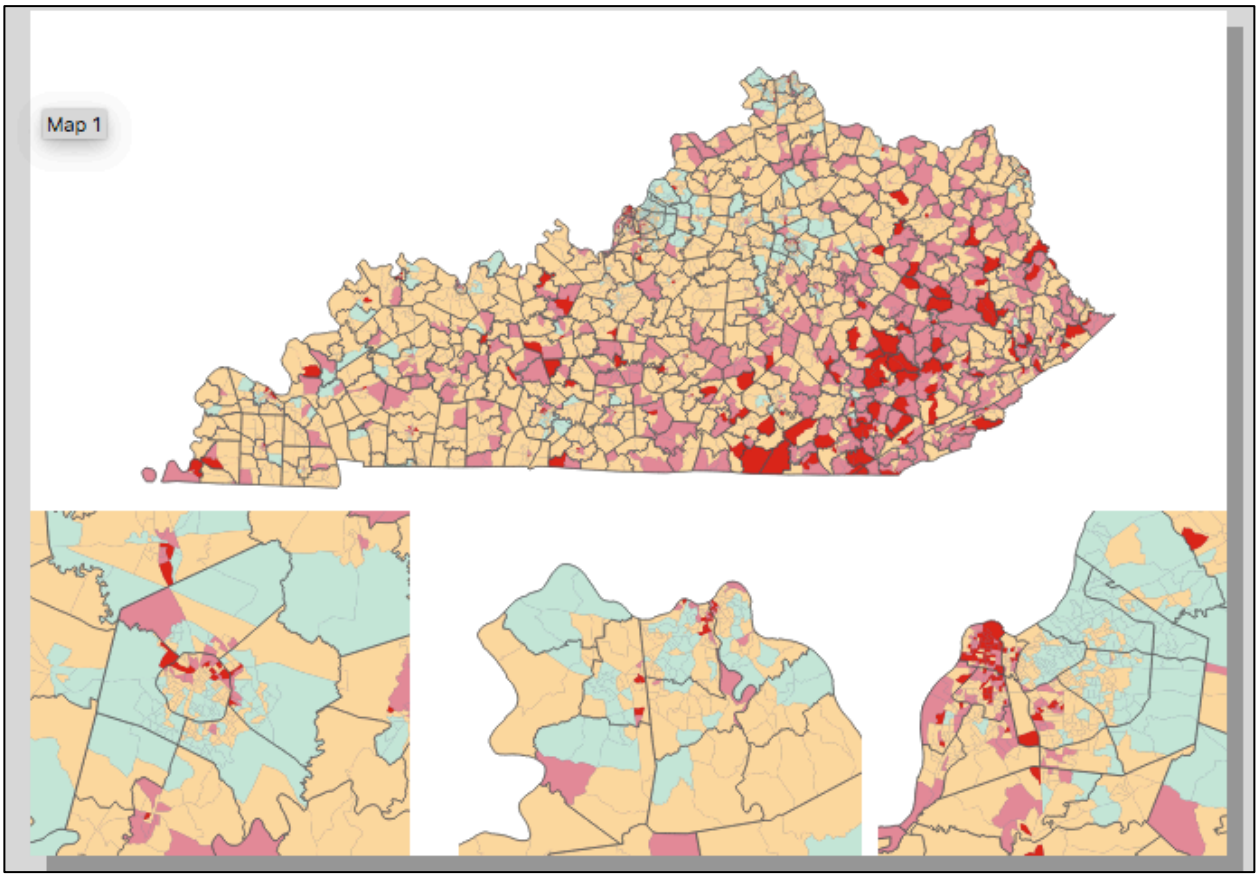
in the “Main Properties” section of Map 1; I recommend 3200000 - 2800000) until the print layout resembles the screenshot to the right:

Now, as we discussed, the three major cities of Lexington, Louisville, and Covington are very tough to see at this scale. To account for this, we will add three **inset maps** to the canvas. This can be done by 1) adjusting the scale in our QGIS data frame, and 2) adding a new “map” to the print layout.

First, zoom to the city of Lexington in the main QGIS data frame. Then, by clicking the **Add new map** button, add a new map to the bottom-left hand corner of the canvas, like so:

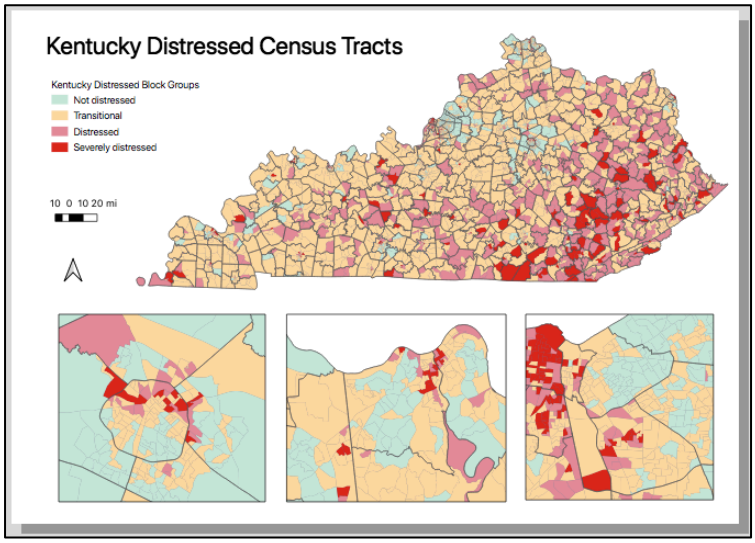


Complete this for Louisville and Covington until you have all three cities included as inset maps in your print layout. See the below screenshot for reference, and be aware that you may have to continue playing around with scale/do some laying out on your own to get the margins right. I recommend resetting the scale for your inset maps to around 300000 each.



Our final map will have the full state shown, with insets for Lexington, Covington, and Louisville. The map should also contain:

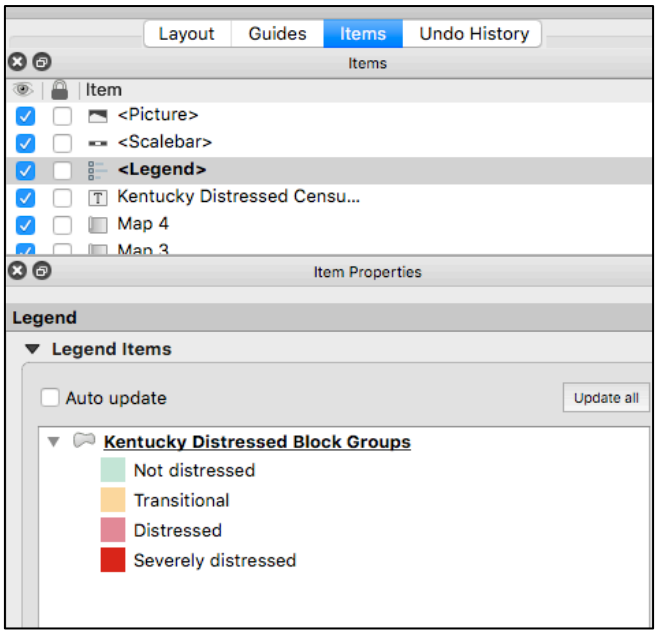
- Legend
- Title
- North arrow
- Scale
- Author's name
- Data source
- Title
- Titles for the city insets



Your final map should **resemble** the one to the right – but keep in mind that this is a **template**, and if you submit a final map that looks exactly like this one you'd be missing some of the key requirements (e.g., Titles for city inserts)!

Things to consider:

- **Styling the legend** can be tricky. Be sure that when you are updating the legend, you have it selected in the **Items** pane, and that **Auto update** is unchecked, as seen in the screenshot to the right.
- **Adding the north arrow** is done by clicking the **Add image button** and then expanding the **Search directories** tab under the image item properties
- You can add shapes and other boxes to the print layout, which may help in organizing the inset maps
- Refer to the documentation for the Print Composer here:



https://docs.qgis.org/2.8/en/docs/user_manual/print_composer/print_composer.html

It can be extremely helpful in case you get lost trying to figure out which button does what!

STEP 4: REVIEW & SUBMIT

You should submit completed materials for Lab 6 via Canvas by Monday, 3/16 at 11:59pm. The completed materials include:

- A Word document containing answers to the questions posed throughout the lab
- One map, formatted in print composer w/ all requisite elements, in .pdf or .png

Be in touch if you have any questions!

For extra credit worth 1 point, add the rural/urban map to your layout. It may resemble the following before your export: