GIS III: GIS Analysis Toolset
Mapping Uncertainty Exercise

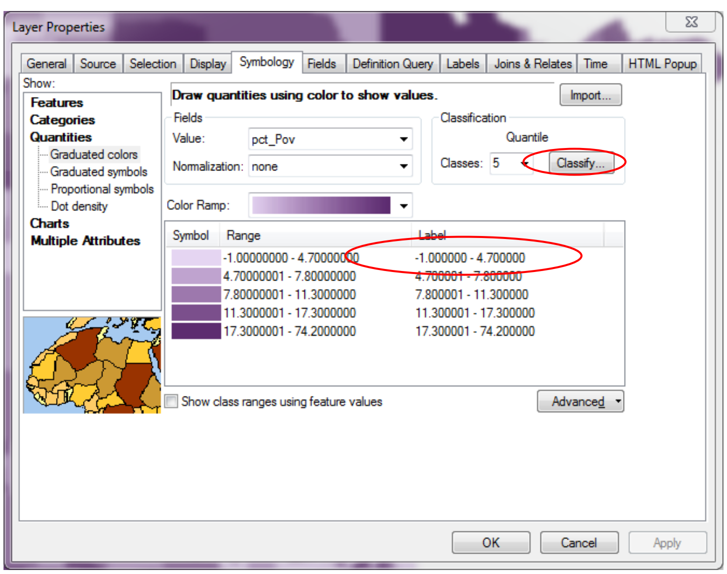
*** Files needed for exercise: *MN_Tracts_ACS_2015_5yr.shp, MN_county10_prj_carto.shp, MN_HrtMrt_65p_05_07.dbf, MN_HrtMrt_65p_13_15.dbf*

Goals: The goal for this exercise is to explore techniques to map and evaluate uncertainty in data estimates.

Skills: After completing this exercise, you will be able to map error measurements using overlay techniques, and check for significant differences among values and classes. And evaluate the statistically significant value difference over time.

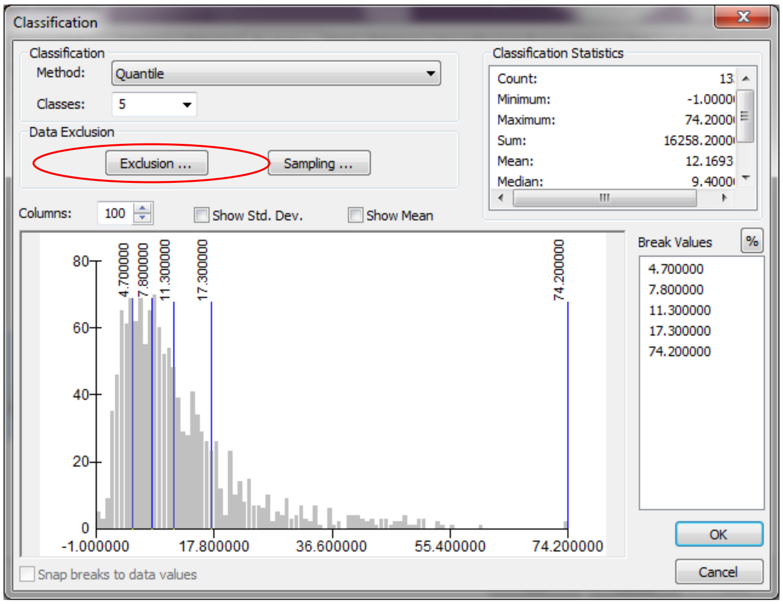
Creating Overlay Maps

1. Open a new **Blank Map** in ArcMap and add *MN_Tracts_ACS_2015_5yr.shp.* Open the attribute table. These data come from the 2015 American Community Survey 5-year estimates for Minnesota. This table has estimates for percent below poverty along with corresponding 90% margins of error (MOE) expressed.
2. Go to **Layer Properties** and symbolize the layer based on **pct_Pov** using a quintile classification scheme and appropriate color ramp.
3. You should note that there are negative values. We do not want to include these in our classification scheme since they represent missing values.



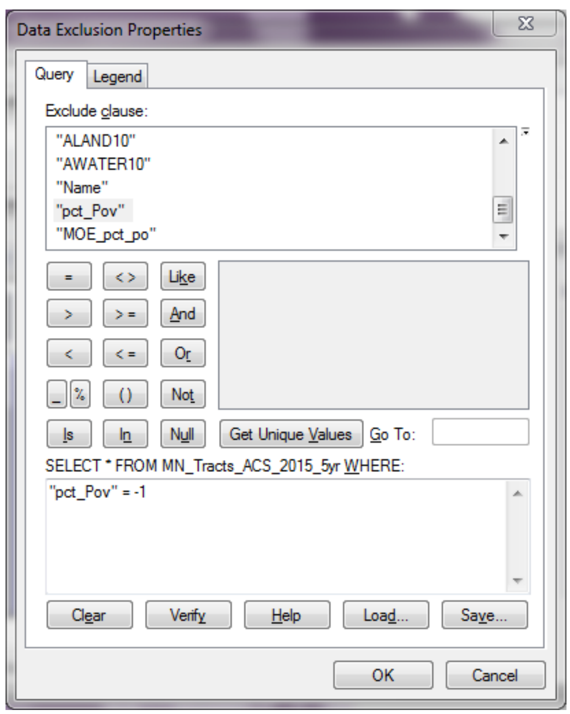
4. To address these negative values that represent NODATA select the *Classify* tab and then the *Exclusion* tab.

GIS III: GIS Analysis Toolset
Mapping Uncertainty Exercise

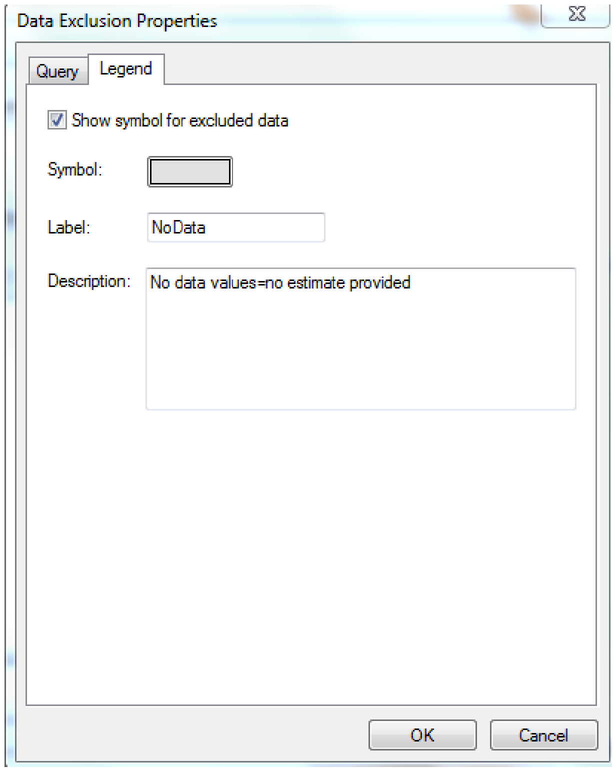


5. On the Query tab build the following query:

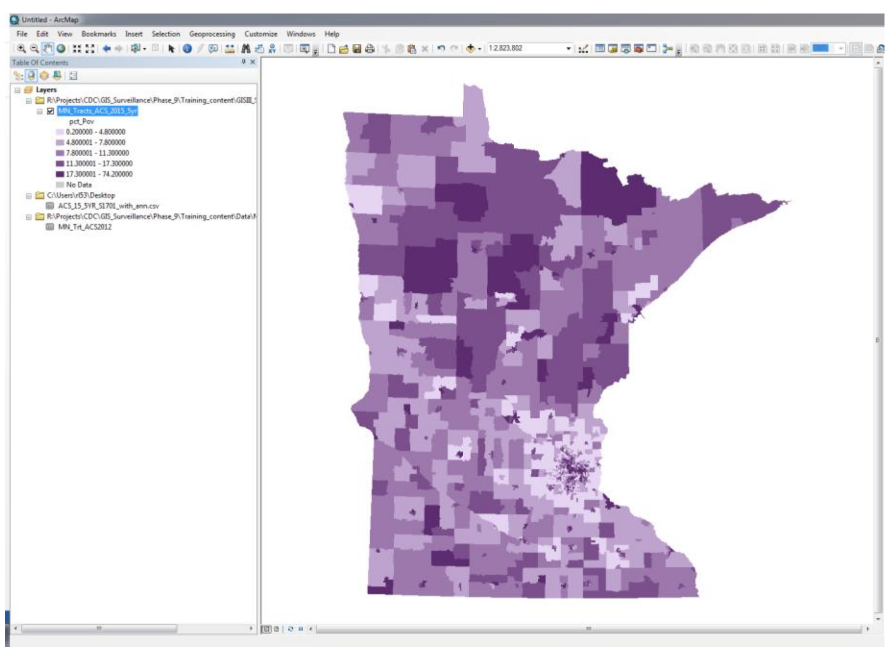
   " pct_Pov" = -1. This will exclude pct_Pov values that equal to -1 (no data)



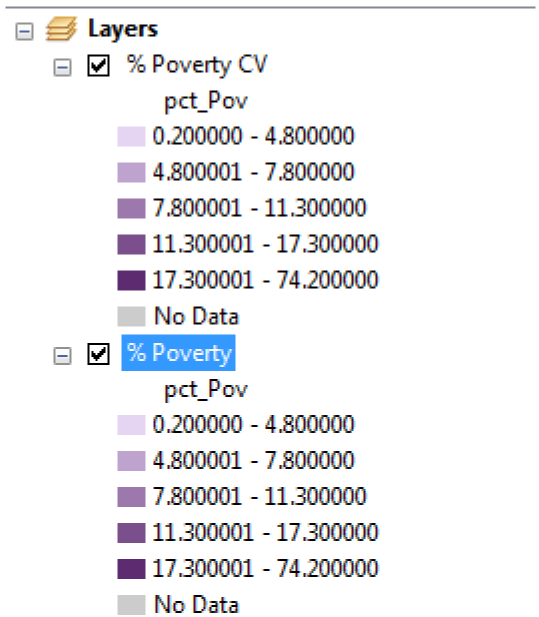
Next click on the **Legend** tab. After checking the **Show symbol for excluded data** check box, pick a symbol and provide a label and brief description.



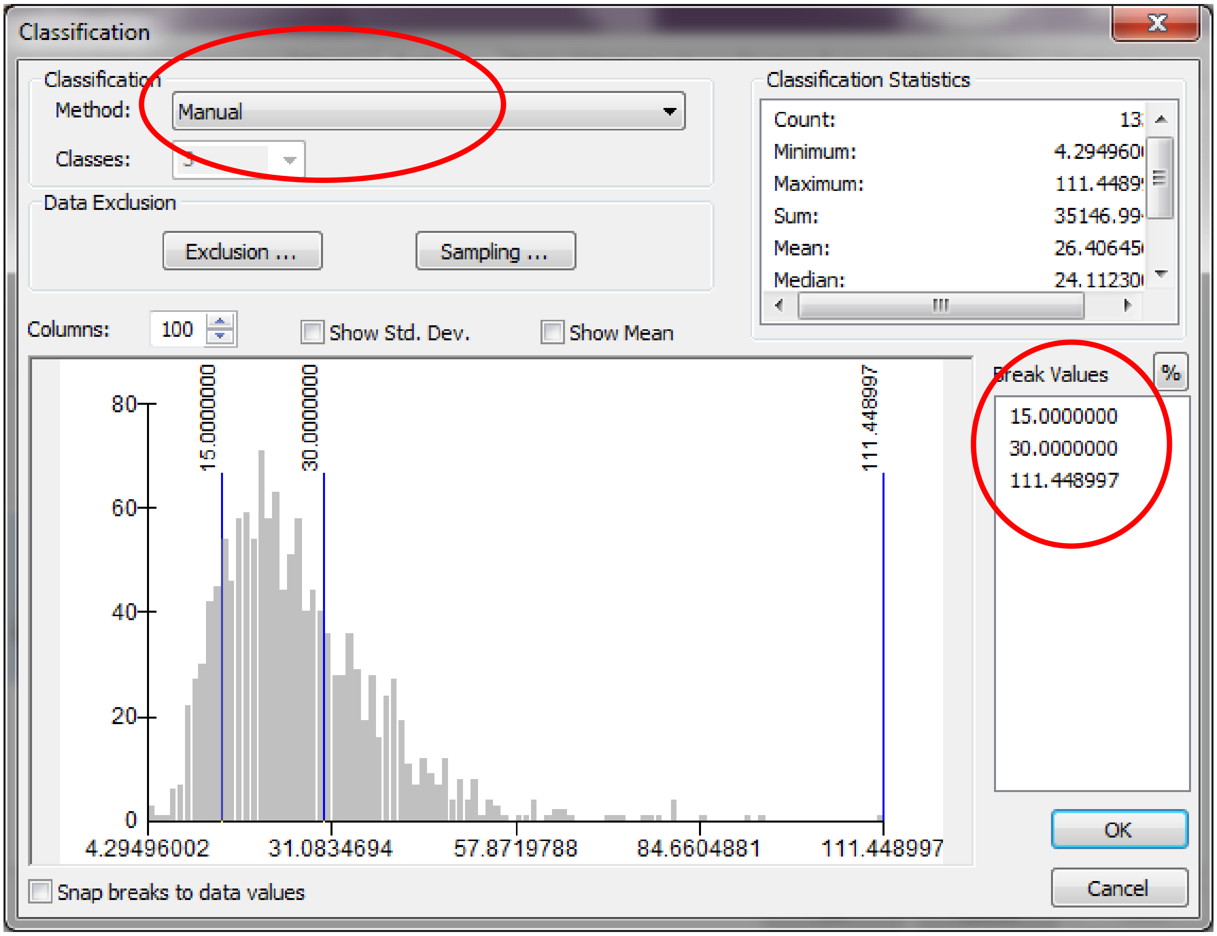
Take a look at the result: the -1 values are no longer part of the distribution and do not affect your classification scheme.



6. Make a copy of your *MN_Tracts_ACS_2015_5yr*. Right click on the layer and select **Copy**, then right click on the Data Frame “Layers” and select **Paste**. Rename the first layer % Poverty CV and the second layer % Poverty.

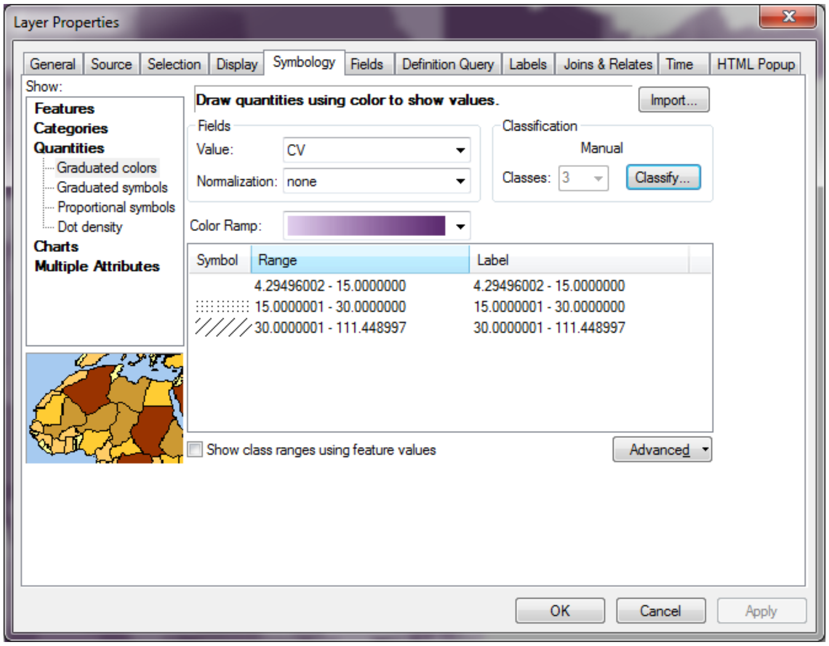


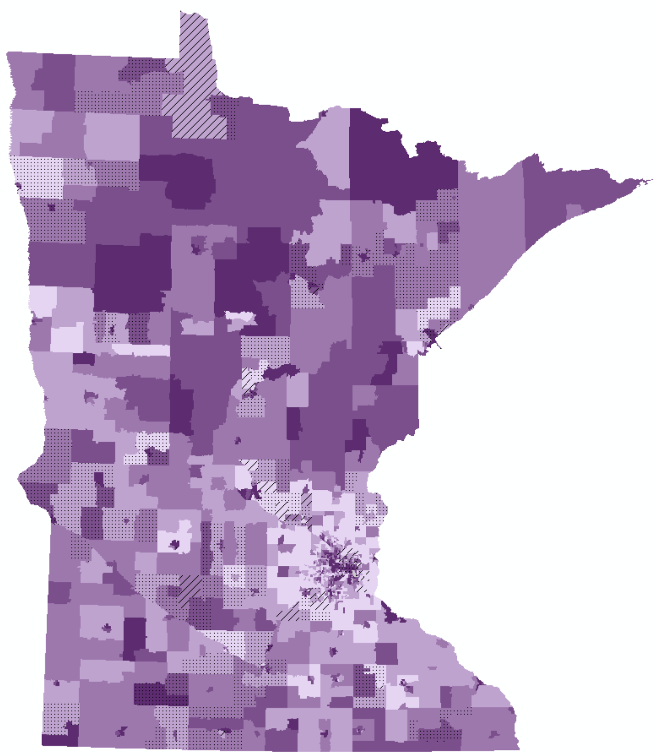
8. We will derive Coefficient of Variation from the table now. Right click to **Open Attribute Table** for *% Poverty CV layer*.
9. **Add Field** *CV* as float in your table. Use the following equation in the **Field Calculator**: [MOE_pct_po]/ [pct_Pov]/1.645 *100
10. For the CV layer, you will symbolize the CV levels that will overlay on top of the poverty values. Go into the **Layer Properties**, **Symbology** tab. Choose **Quantities** > **Graduated colors**. Select the **CV** field.
11. Click the **Classify** button. CV values range from 4.3 to 111.45, with a few outliers. There are no hard and fast rules for classifying these values – U.S. Census case studies suggests the following categories: High reliability: CVs less than 15%; Medium Reliability: CVs between 15-30% - be careful; Low Reliability: CVs over 30% - use with extreme caution. You should ideally choose no more than three classes. Adjust the number of classes appropriately, and select **Manual** as the method. You can then type in values into the **break values box** to set your thresholds. Click **OK**.



12. You can now adjust the fill pattern for your classes of CV. Double click on each symbol swatch to change the fill pattern. A common technique is to have the lowest class represented as empty/hollow and other classes with a hatch or dot fill. When finished, click **OK**.

GIS III: GIS Analysis Toolset
Mapping Uncertainty Exercise

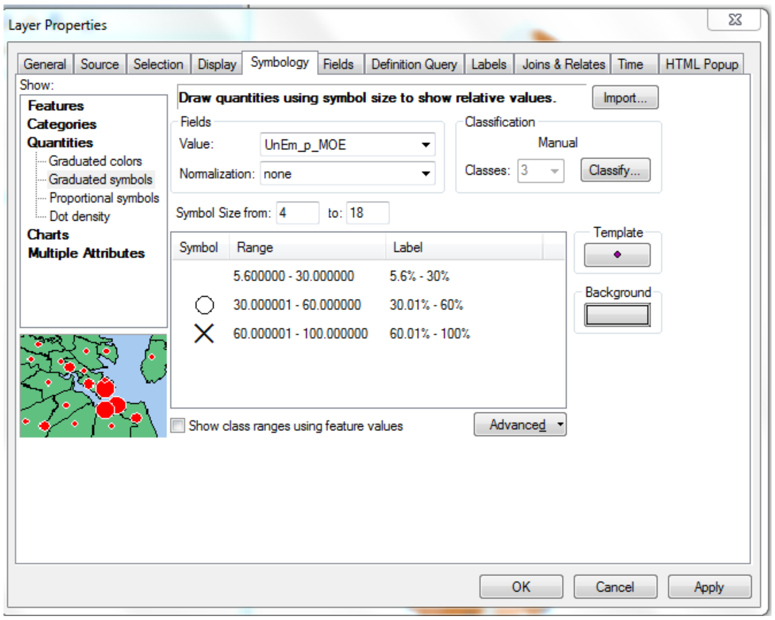


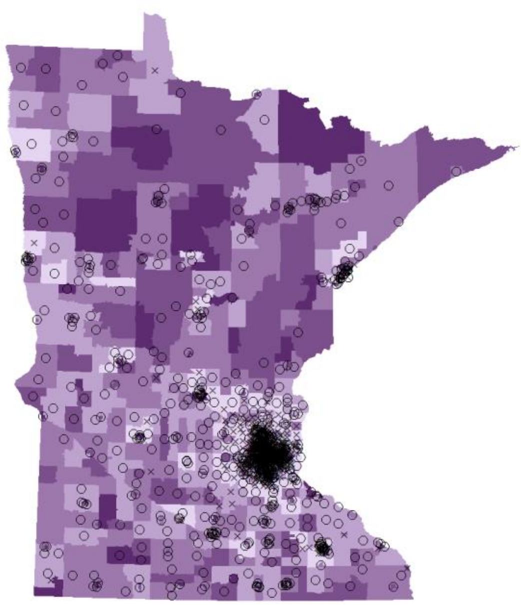


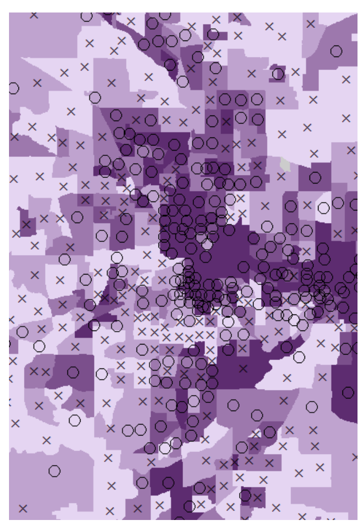
13. Examine the legibility of your map. Some color ramps work better with patterns than others. Alternatively, you could create a second data frame and display two maps side-by-side, one with the percent poverty rate and the other with the CVs.

GIS III: GIS Analysis Toolset
Mapping Uncertainty Exercise

14. Another approach is to use symbols to distinguish between classes of CV. For your CV layer, go to Layer Properties. In the Symbology tab, select **Quantities** > **Graduated Symbols** and choose **CV** as the Value. Use the same classification scheme developed above. Change the symbols so that the lowest class is empty, the second class is a circle, and the third is an 'X'. Set the **Background** to hollow. Click **OK** and evaluate your map for legibility.





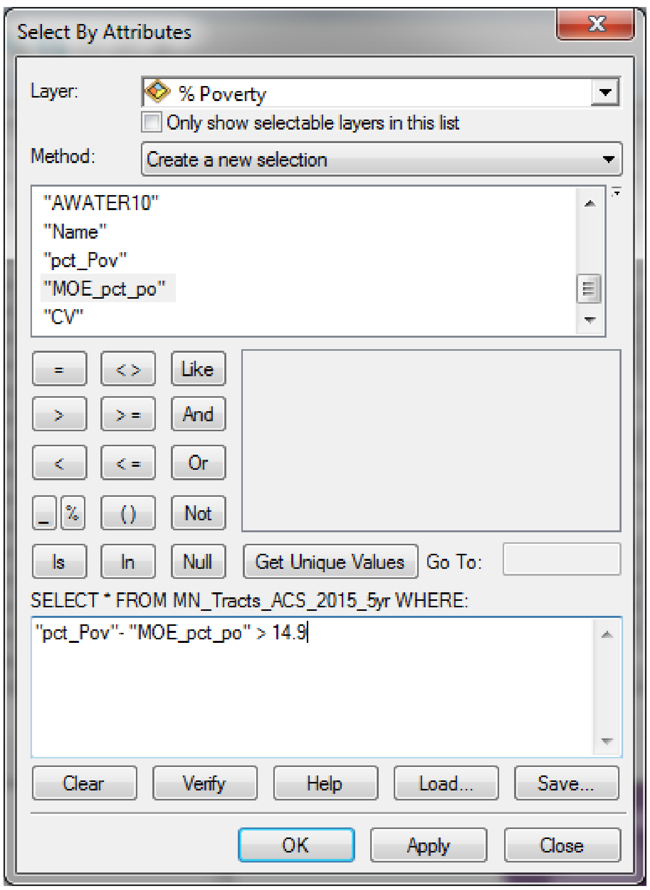


What do you think? Which of the two techniques work best for this case?

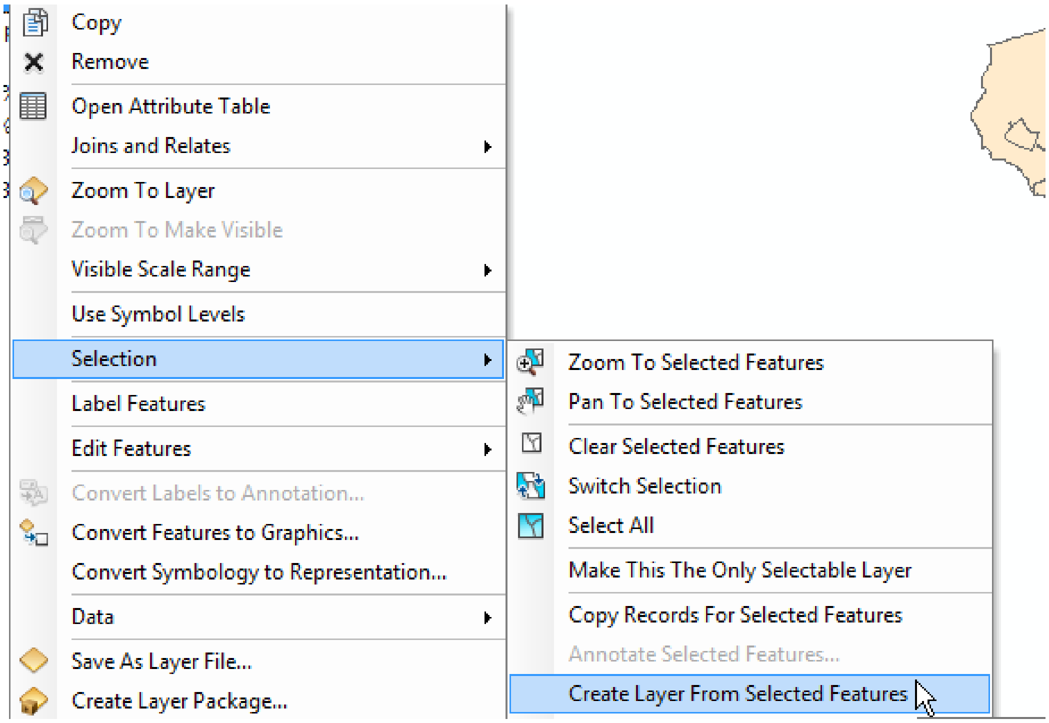
GIS III: GIS Analysis Toolset
Mapping Uncertainty Exercise

Evaluating Significant Difference

1. To test for statistically significant differences between a fixed value (state average, a specific estimate etc.) and your estimates, you can write selection queries that make use of MOE or confidence intervals. Remember that your Estimate ± MOE gives you your confidence interval bounds in most cases (the exceptions are CDC Wonder and Interactive Atlas data which use an alternative statistical technique and do not report MOE).
2. Let's compare % poverty for MN Census Tracts in 2015 to the National value for poverty rate that same year: **14.9%**
3. You will select tracts that are significantly different than this value. From the **Selection** menu, choose **Select by Attributes.** Choose the **% Poverty** layer.
4. Build a query: ***"pct_Pov"- "MOE_pct_po" >14.9***. This will select tracts that are significantly higher than the national value for the poverty rate in 2015 (i.e. the lower bound of the confidence interval is greater than the average). Click **OK**.



5. Right click on your layer and go to **Selection > Create Layer from Selected Features** to create a new layer with your significantly higher tracts.

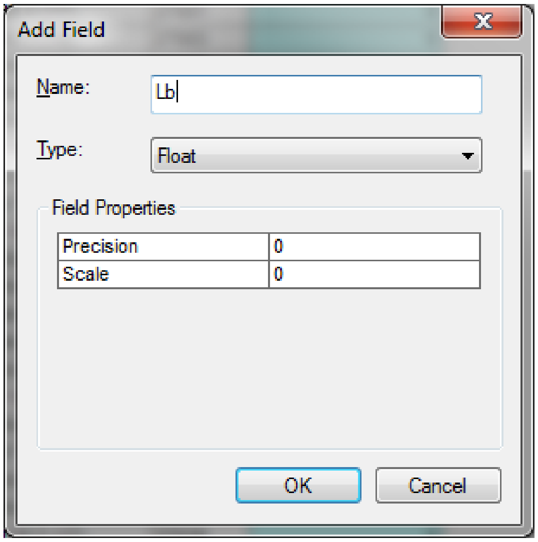


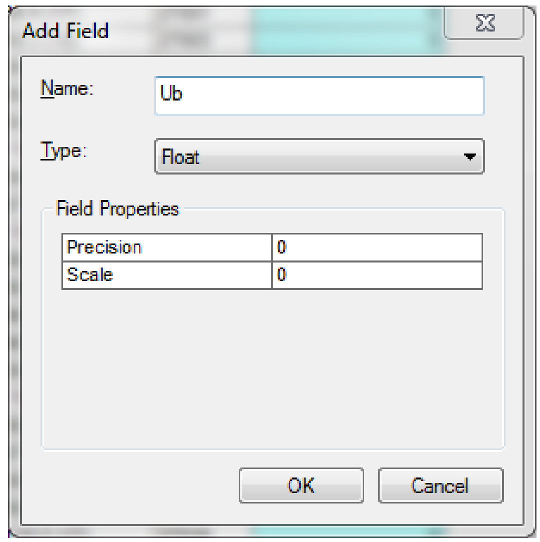
6. Rename your new layer "**significantly higher**." You can now symbolize this layer with a cross hatch or other pattern. Clear your selected features.
7. Build another query: *"pct_Pov"+ "MOE_pct_po" < 14.9*. This will select tracts that are significantly lower than the national value for poverty rate (i.e. the upper bound of the confidence interval is less than the average). Click **OK**. Create a new layer and symbolize and name appropriately.
8. Use the techniques you just learnt to display this information.
9. Think about how other queries could be written. Upper and lower confidence bounds could be used instead of "estimate ± MOE." You could also use upper and lower confidence bounds instead of the fixed value, in which case you'd want to see if Estimate ± MOE was > the upper bound (significantly higher) or < lower bound (significantly lower).

Change: Evaluating Significance over Time

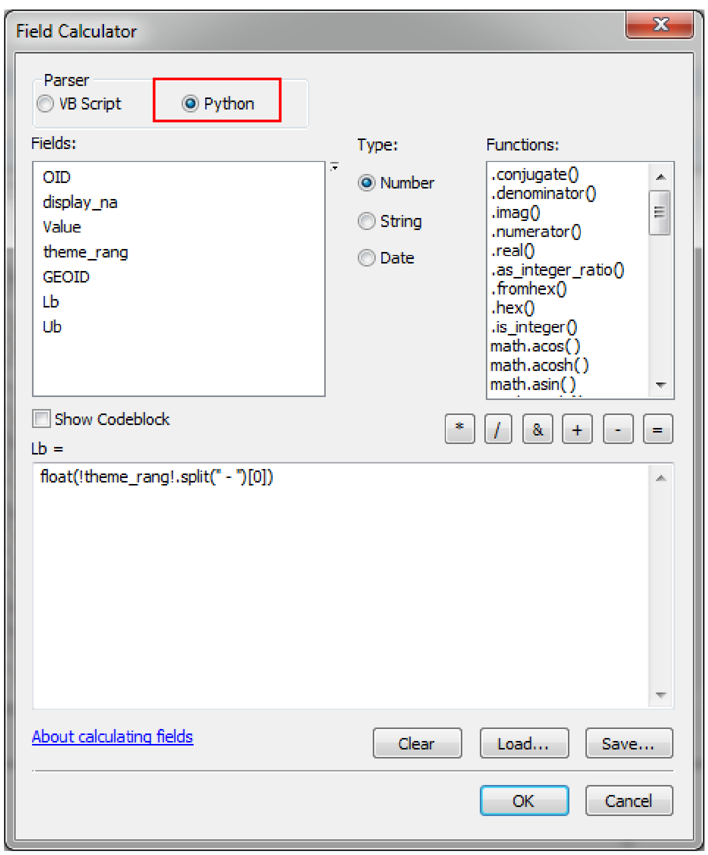
1. Add *MN_HrtMrt_65p_05_07.dbf* and *MN_HrtMrt_65p_13_15.dbf* to the workspace. Right click to **Open Attribute table** for *MN_HrtMrt_65p_05_07.*
2. In order to calculate the significant change, we need the upper and lower bounds for both estimates to implement the condition equation: $|E_1 – E_2| > (Upper_1 – Lower_1) + (Upper_2 – Lower_2)$. However, the confidence interval in the dataset looks like this after downloaded from CDC's Interactive Atlas: 769.4 – 836.4 (17). We will extract both upper boundary and lower boundary for

the calculation. Use **Add Field** to add two new float field to the table and name them as **Lb** and **Ub**.





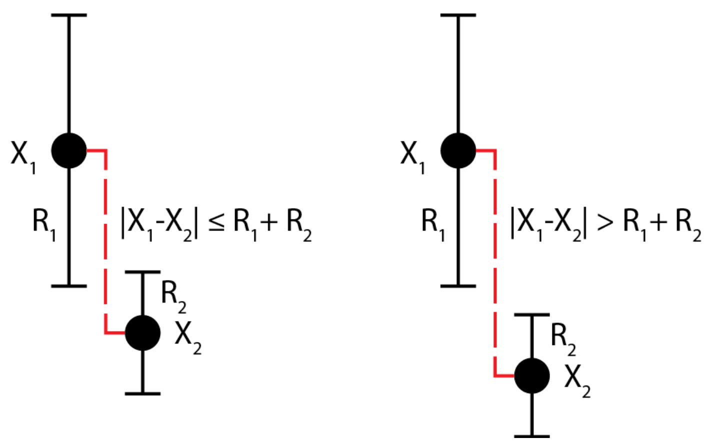
3. Right click on **Lb** and select **Field Calculator**. Make sure you check **Python** instead of ~~VB Script~~. Put **!theme_rang!.split(' - ')[0]** in the equation box and click **OK**.



4. Similarly, calculate the field for **Ub** by using this equation: **float(!theme_rang!.split(" - ")[1].split(' (')[0])**.
5. Now you calculated the upper bound and lower bound for the confidence interval for *MN_HrtMrt_65p_05_07*. To make our time more productive, we calculated the same variables for *MN_HrtMrt_65p_13_15*. You can start to evaluate the change of the heart disease mortality rate for population over 65 years old between 2006 and 2014.
6. Add *MN_county10_prj_carto.shp* to the workspace. **Join** both *MN_HrtMrt_65p_05_07 and MN_HrtMrt_65p_13_15* to *MN_county10_prj_carto*.
7. Use **Add Field** to add 2 new field for *MN_county10_prj_carto*: Diff - float, and sig - Short integer.

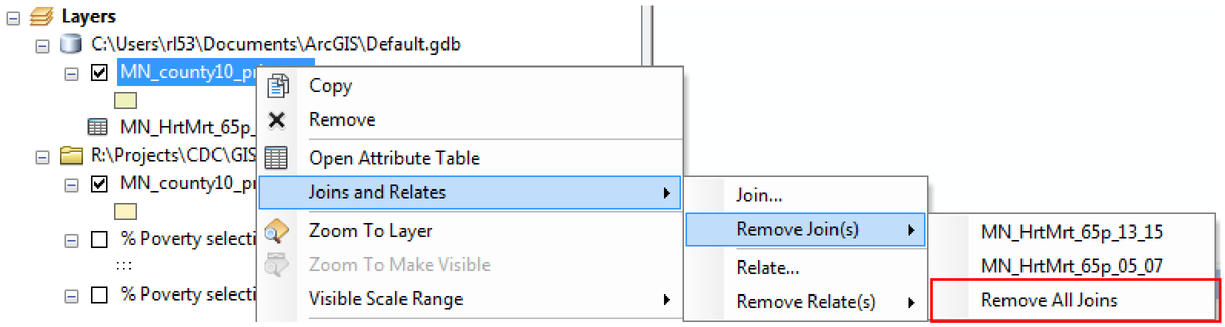
MN_county10_prj_carto.Diff	MN_county10_prj_carto.sig

8. Right click *MN_county10_prj_carto.Diff* to select **Field Calculator**. Make sure you select **Python** as your **Parser** language. Use this equation: **!MN_HrtMrt_65p_13_15.Value! - !MN_HrtMrt_65p_05_07.Value!** to calculate estimate change through years.
9. Right click *MN_county10_prj_carto.sigf* to select **Field Calculator**. Make sure you select **Python** as your **Parser** language. Use this equation: **int(abs((!MN_HrtMrt_65p_13_15.Ub! + !MN_HrtMrt_65p_13_15.Lb!) - (!MN_HrtMrt_65p_05_07.Ub! + !MN_HrtMrt_65p_05_07.Lb!))> (!MN_HrtMrt_65p_13_15.Ub! - !MN_HrtMrt_65p_13_15.Lb!) + (!MN_HrtMrt_65p_05_07.Ub! - !MN_HrtMrt_65p_05_07.Lb!))** . It looks like a complex equation, but what it does is actually evaluating the distance between the mid-range points and the range size as in the following graphic shown.



10. Right click on *MN_county10_prj_carto*, go to **Joins and Relates** tab. Click **Remove All Joins** under **Remove Join(s)** sub-menu.

GIS III: GIS Analysis Toolset
Mapping Uncertainty Exercise



11. Now you can symbolize your difference map by suppressing the insignificant changes. Which method will you use to display the insignificance of the differences? Why?