

*** 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

- 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.

Layer Properties			23
General Source Select	ion Display	Symbology Fields Definition Query Labels Joins & Relates Time	e HTML Popup
Show: Features Categories	Draw quar Fields	ntities using color to show values. Import	
Quantities Graduated colors Graduated symbols	Value: Normalizatio	pct_Pov Quantile Classes: 5 Classify	\supset
Dot density	Color Ramp:	■ 100 000 000 000 000 ▼	
Multiple Attributes	Symbol R -1 4. 7. 11 17	tange Label 00000000 - 4.7000000 -1.000000 - 4.700000 7000001 - 7.800000 4.700001 - 7.800000 8000001 - 11.300000 1.300001 - 11.300000 1.3000001 - 17.300000 11.300000 7.3000001 - 74.200000 17.300001 - 74.200000 1s ranges using feature values Advances	1
		OK Cancel	Apply

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



Classification	_	×
Classification	Classification Statistics	
Method: Quantile	Count:	13. 🔺
Classes: 5 👻	Minimum:	-1.0000
Data Exclusion	Maximum:	74.2000 =
Euclosian Canalian	Sum: Mean:	12 1693
Exclusion Sampling	Median:	9,4000
	۰ III	4
Columns: 100 Show Std. Dev. Show Mean		
	8	Break Values 🧏 %
	50	4.700000
4 2 4 2	74	7.800000
60+		17 300000
		74,200000
40-+		
20-		
a ball a state of the ball of the second second		
-1.000000 17.800000 36.600000 55.400000	74.200000	ОК
Snap breaks to data values		Cancel

5. On the Query tab build the following query:

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

Data Exclusion Properties	23
Query Legend	
Exclude <u>clause</u> :	
"ALAND10"	* -
"AWATER10"	
"pct_Pov"	=
"MOE_pct_po"	-
= <> Li <u>k</u> e	
> >= And	
< <= O <u>r</u>	
_% () Not	
Is In Null Get Unique Values Go To:	
SELECT * FROM MN_Tracts_ACS_2015_5yr WHERE:	
"pct_Pov" = -1	*
	_
Load Save	
OK Car	ncel

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.



	ata Exclusion A	Properties	23
	Query Leger	nd	
1	Show sym	bol for excluded data	
	Symbol:		
	Label:	NoData	
	Description:	No data values=no estimate provided	
		ОК	Cancel
			Cancer

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





 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.
- Add Field CV as float in your table. Use the following equation in the Field Calculator: [MOE_pct_po]/ [pct_Pov]/1.645 *100
- 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**.



General Source Sele	ction Disp	ilay S	Symbology	Fields	Definition Qu	ery Lab	els Joins & Re	lates Time	HTML Popus
Show: Features Categories Quantities 	Draw Fields Value: Normal	quanti zation:	ties using CV none) color (to show valu	Class	sification Manual ses: 3 v	Import	
Charts Multiple Attributes	Color Ra Symbol	mp: Rar 4.29	nge 0496002 - 1 0000001 - 3	5.00000 0.00000		Label 4.294960 15.00000	02 - 15.000000 01 - 30.0000000		
	Show	class i	ranges usin	g feature	values			Advanced	T



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.



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.

General	Source	Selection	Display	Symbology	Fields	Definition Query	Labels	Joins & Relates	Time	HTML Popup
how:										
Feature	es		raw qua	ntities usin	g symbo	l size to show r	elative v	alues.	port	
Catego	ries	_ F	ields				Classifica	tion		
Quanti	ties	V	alue:	UnEm_p	_MOE	•		Manual		
Grad	duated col duated syn	ors nbols N	lormalizati	on: none		-	Classes:	3 v Clas	sify	
Prop Dot	ortional sy density	mbols Sy	mbol Size	e from: 4	to: 18					
Charts Multiple	e Attribu	tes S	ymbol	Range		Label		Ten	oplate ♦	
			5	5.600000 - 30.	000000	5.6% - 30%				
			\sim	0 00001 6	000000	20.01% 00%		Back	ground	
			\mathcal{O}	0.000001-0	000000	30.01%-00%	•			
Ŕ		\mathbf{i}	X	60.000001 - 10	0.00000	0 60.01% - 100	%			
Ÿ	्रे	× [Show cla	ss ranges usir	ng feature	values	Adva	ance <u>d</u> •		



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



Evaluating Significant Difference

- 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.
- 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.

Select By Attributes	×
Layer: 🔷 % Poverty	•
Method: Create a new selection	-
"AWATER10" "Name" "pct_Pov" "MOE_pct_po" "CV"	× 11
= <> Like > > And <	
SELECT * FROM MN_Tracts_ACS_2015_5yr WHERE:	
"pct_Pov"- "MOE_pct_po" > 14.9	*
Clear Verify Help Load Sar	ve
OK Apply Cla	ose

 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.
- 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).</p>

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.*
- In order to calculate the significant change, we need the upper and lower bounds for both estimates to implement the condition equation: |E₁ E₂| > (Upper₁ Lower₁) + (Upper₂ Lower₂). 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**.

Add Field	Add Field
Name:	Name: Ub
Type: Float	Type: Float
Field Properties	Field Properties
Precision 0	Precision 0
Scale	Scale
OK Cancel	OK Cancel

 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.

Field Calculator	X
Parser VB Script Python Fields: OID display_na Value theme_rang GEOID Lb Ub	Type: Functions: <pre> Number .conjugate() .denominator() .imag() .numerator() .real() .as_integer_ratio() .fromhex() .hex() .is_integer() math.acos() math.acos() math.acos() </pre>
Show Codeblock	* / & + - =
float(!theme_rang!.split(" - ")[0])	A
About calculating fields	Clear Load Save
	OK Cancel



- 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	
----------------------------	---------------------------	--

- 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.



🖃 🥩 Layers				
🖃 🧊 C:\Users\rl53\Documen	ts\ArcGIS\Default.gdb			
□ 🗹 MN_county10_p	ђ Сору			
MN_HrtMrt 65p	Remove			
🖃 🚞 R:\Projects\CDC\GIS	Open Attribute Table			
□ 🗹 MN_county10_p	Joins and Relates	•	Join	
🖃 🔲 % Poverty selecti 🔍	Zoom To Layer		Remove Join(s) MN_HrtMrt_65p_13_15	
	Zoom To Make Visible		Relate MN_HrtMrt_65p_05_07	
🖃 🔲 % Poverty selecti	Visible Scale Range	•	Remove Relate(s) Remove All Joins	

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?

