Mapping Uncertainty

GIS III: GIS Analysis Toolset

CEHI
Children's Environmental Health Initiative

Learning objectives

- Understand the concept of confidence for estimates
  - Explore common measures of reliability
  - Evaluate statistical differences between estimates
- Consider how you will communicate this information

Understand uncertainty in data estimates and explore different mapping techniques used to display unreliable data.

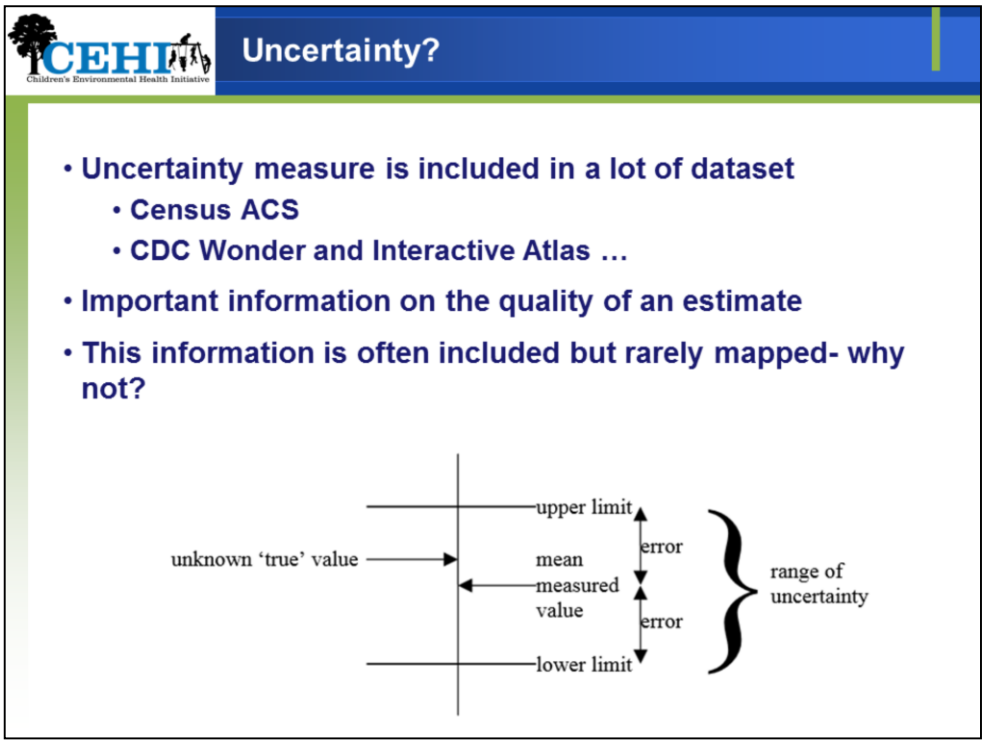
• Choose an appropriate error measurement and define your error threshold

• Compare the distribution and magnitude of error measurements to determine reliability of estimates and to assess which variables are suitable for mapping

• Choose a mapping technique depending on your geography and reader that maximizes legibility

• Consider what statements can be made about statistical significance between values

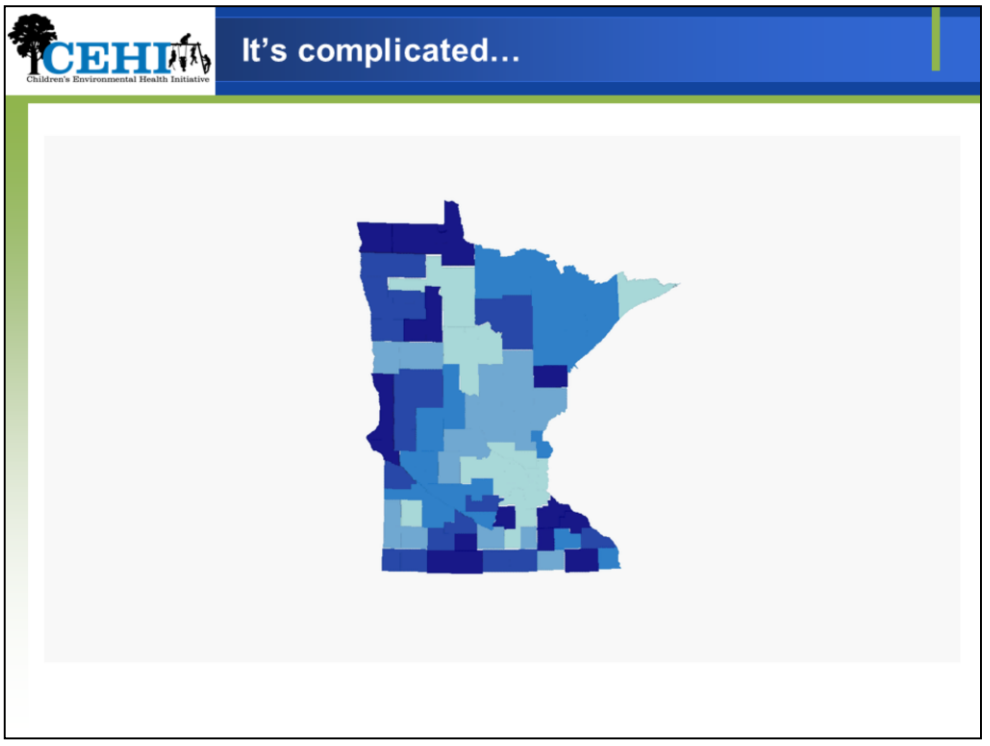
Hunt-Lenox Globe



As of 2010, the American Community Survey (ACS) has replaced the long form decennial census, and resulting changes in the sampling procedures have created errors and uncertainty that must be accounted for when using this data.

Other data sources, such as the CDC's Diabetes Interactive Atlas, also recognize error in their estimates. Data quality information (such as margins of error or confidence intervals) is often included in datasets, but spatial mapping usually overlooks this information. Incorporating data quality information into maps is conceptually and technically complex given that data and data quality measures have to be shown together.

The challenge is
to produce maps that address data quality issues accurately while remaining comprehensible to map readers. There are a variety of techniques being developed to address this.



In ArcScene, you can turn a flattened map into a nearly 3D representation of

values and confidence intervals

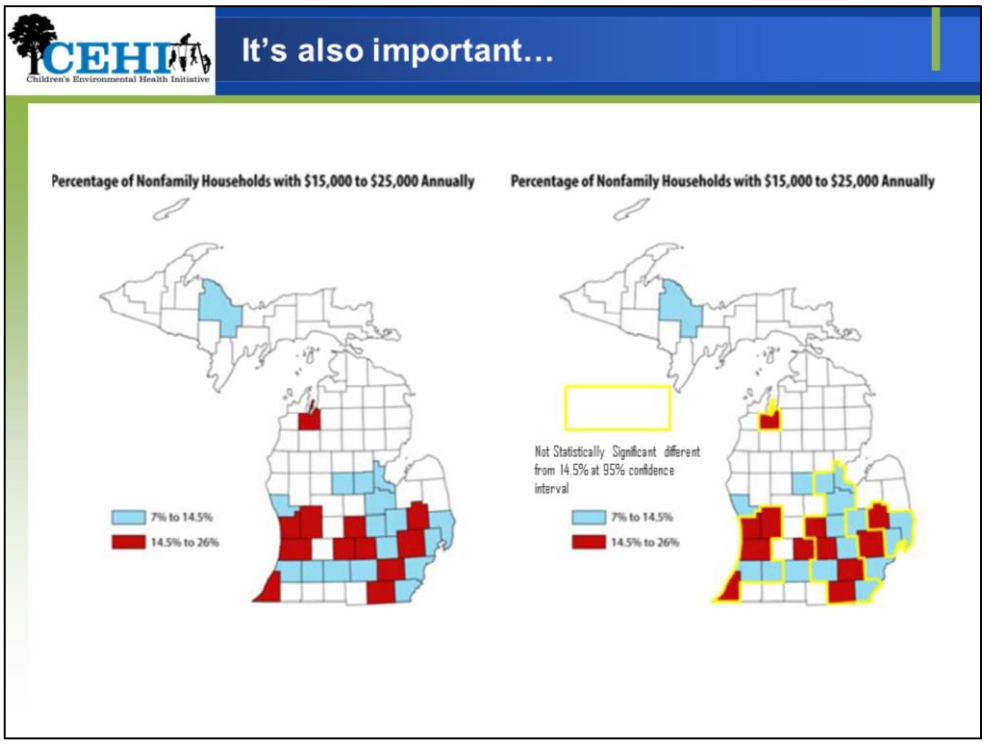
Pros

-Identifies estimates that are statistically significant from another selected

estimates visually on a map in a third dimension

Cons

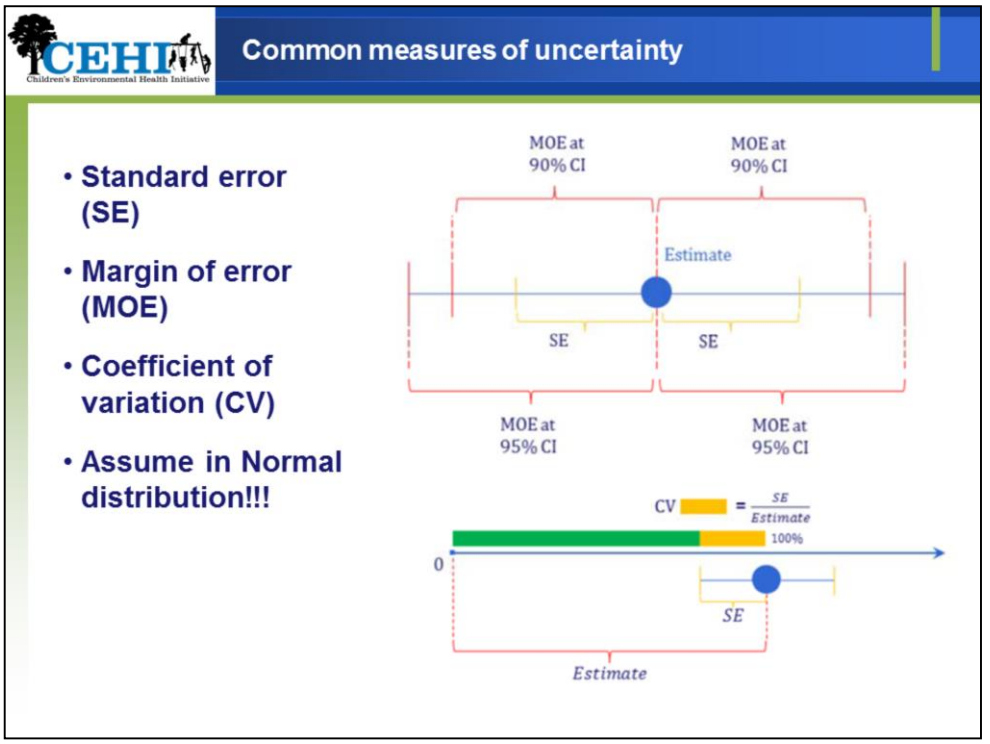
-Challenging to share other than in a dynamic setting



Poverty threshold example: MI 2012 ACS county level

Undermine the credibility of a particular message

Mislead or confuse the target audience



Sampling error – arises when you take a sample. Reflects the difference between the sample and population estimates.

• Standard error (SE) – measure of imprecision in an estimate due to sampling,

standard deviation of the estimate

• Margin of error (MOE) – another measure of sampling error - determined by your

confidence level. The larger the margin of error, the less reliable the estimate. MOE used by the ACS is 1.645 times the standard error (for a 90% confidence level).

• Coefficients of variation (CV) – alternative measure of reliability of an estimate,

computed as standard error/estimate or with ACS data (90% MOE/1.645)/estimate, often multiplied by 100 to express as a percentage.

• To get CV from 90% CI: ((upper limit-lower limit)/(2*1.645*estimate))

• To get CV from 95% CI: ((upper limit-lower limit)/(2*1.960*estimate))

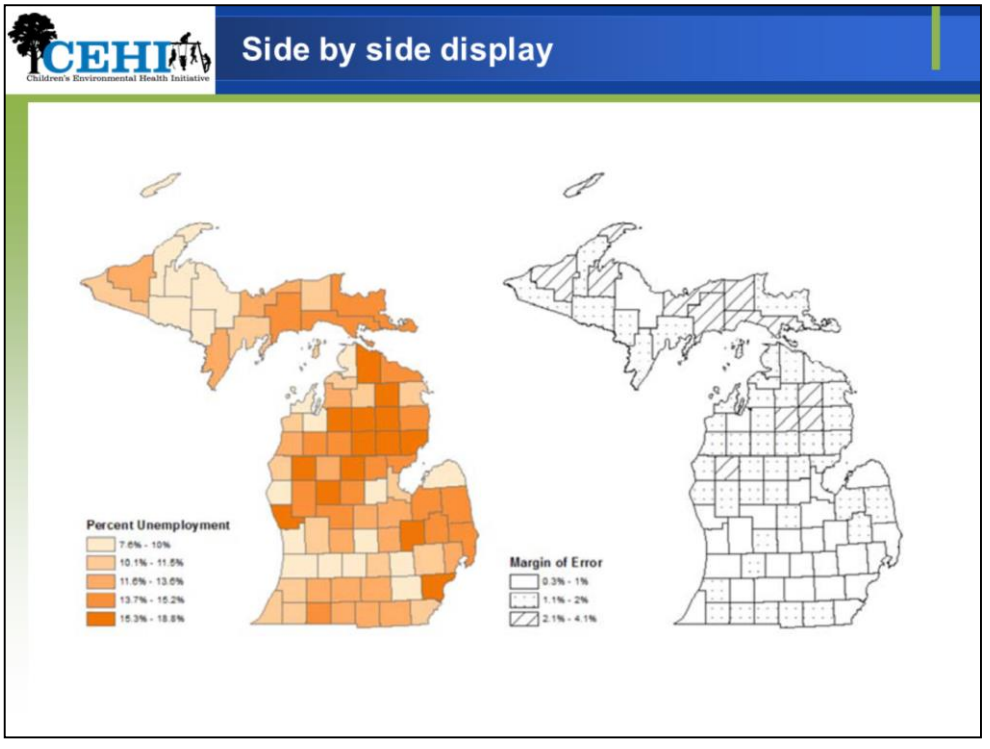
Choosing the appropriate measure to map

choice is not always straightforward. Data on counts/totals/frequencies may be better suited to CV,

Percentages or rates may be better suited to MOE. This depends on the range of your estimated values as well as on the range of your margins of error.

Meaningful information Sometimes rates may simply be too unreliable such that mapping them is

not meaningful or worthwhile



Once you've decided on and calculated a reliability measure, there are several options to

display this information.

Side by side maps are somewhere to start.

Pros

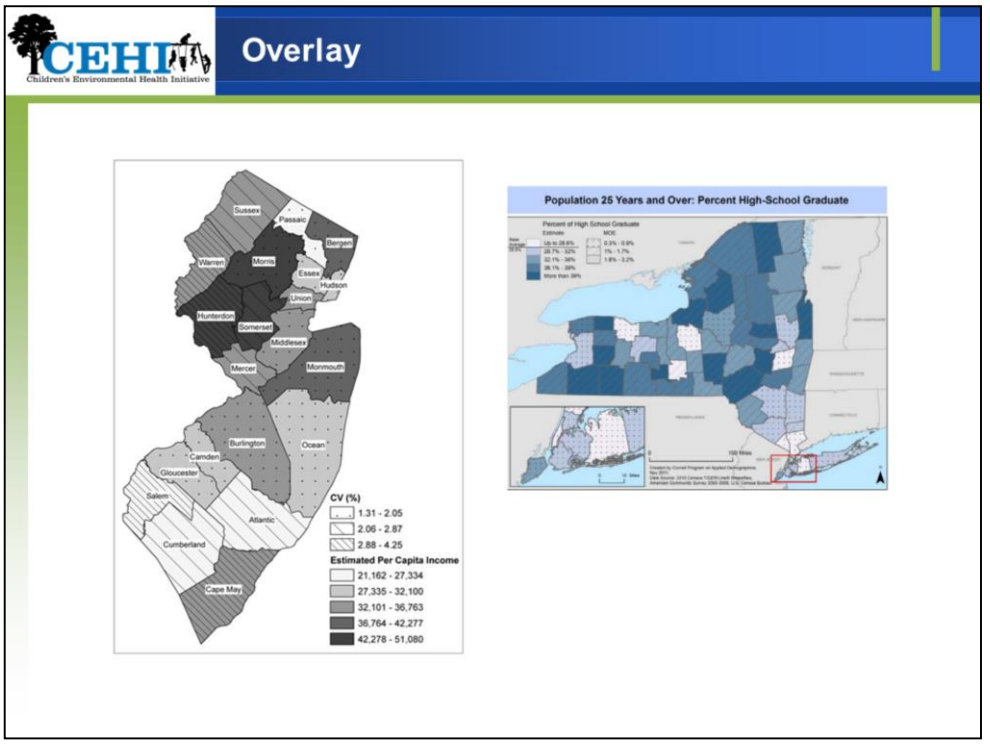
- Displays both the estimates and the reliability measure without cluttering one map/can

easily discern patterns in the estimates and reliability measures in each map individually.

Cons

- Reader must focus back and forth on two maps displays to try and interpret both pieces

of information simultaneously.



Measures of reliability are overlaid directly on top of estimates.

Pros

-Combines information from two side by side maps which is more efficient for readers

-Symbols rather than fill pattern can be less obtrusive

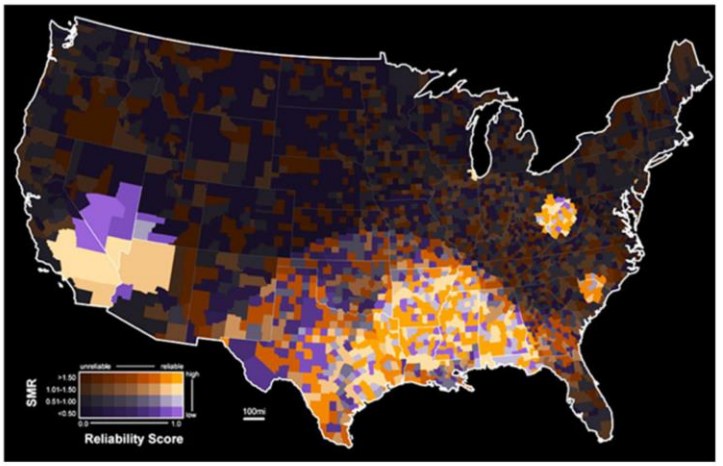
Cons

-With small geographies, any overlay can obscure the underlying variable

-Some color/fill choices can be overpowering or difficult to read

-Does not tell you anything about significance of estimates relative to one another

Value by alpha



Value by alpha maps use a bivariate scale – the colors reflect the values of the primary variable, and the intensity or brightness of the color reflects the reliability (brighter = more reliable).

Pros

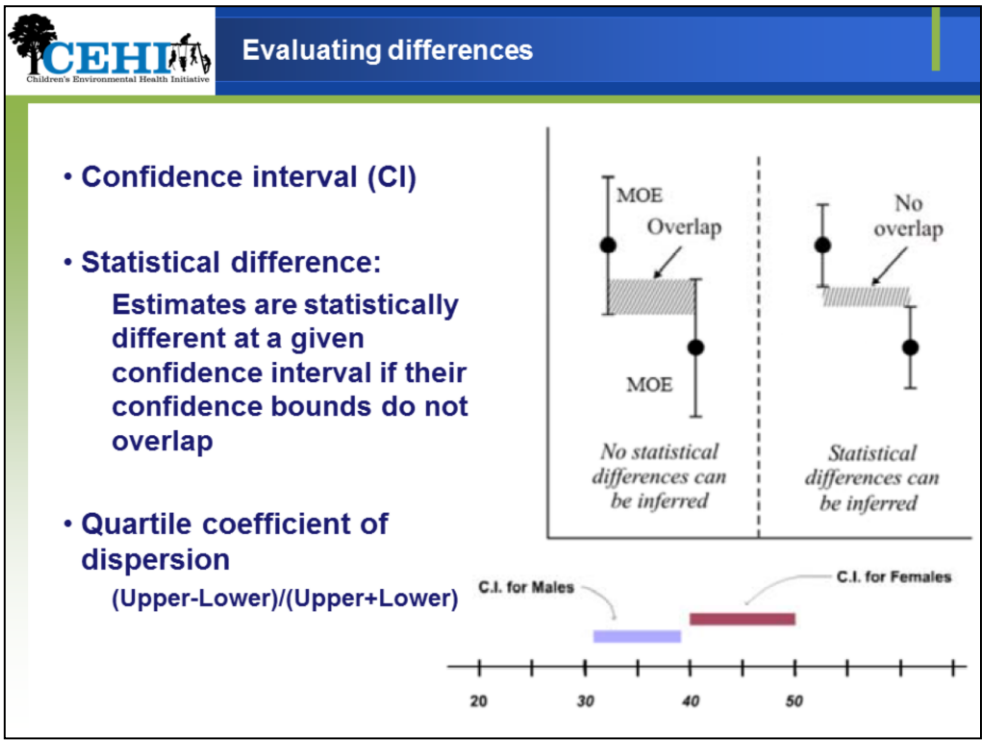
-If done well, more reliable areas are readily apparent or “pop”

-Relies on color/shading rather than patterns which can be distracting

Cons

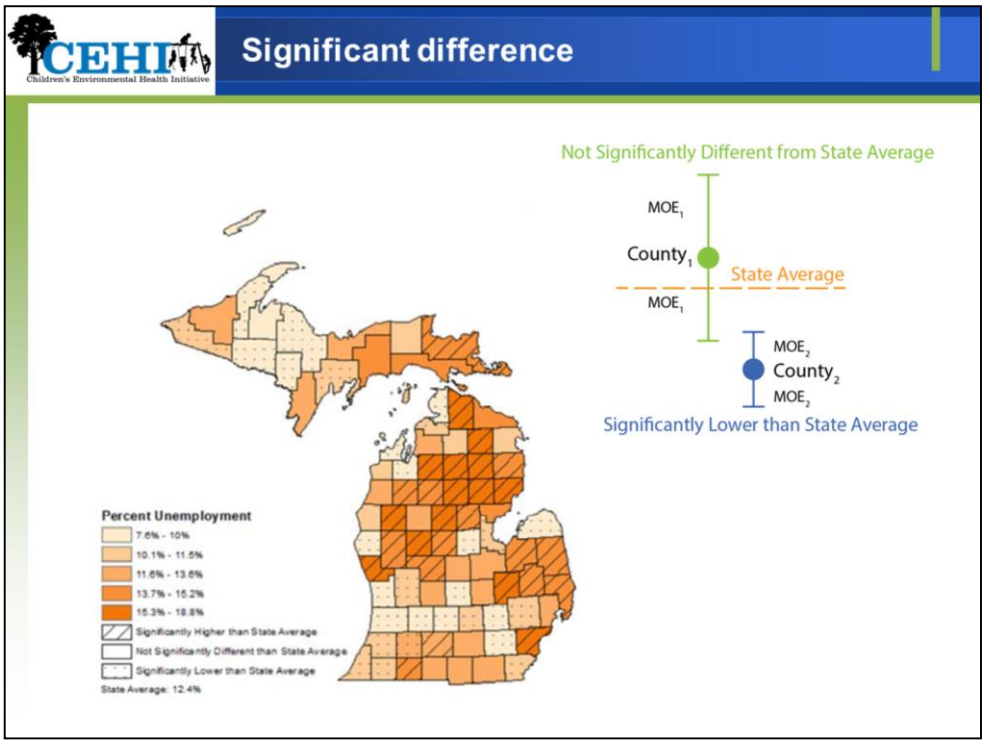
-Challenging for readers to interpret the legend, especially with diverging ramps

-Challenging to produce (require color and use of graphics software)



Confidence interval – bounding limits of where you expect the true estimate to lie, vary depending on the level of confidence you choose (90, 95, 99 etc), calculated as estimate ± MOE

Statistical difference: $|x1 – x2| > MOE1 + MOE2$



Confidence interval (CI)- bounding limits of where you expect the true estimate to lie, vary

depending on the level of confidence you choose (90, 95, 99 etc.), calculated as estimate ±

MOE.

Statistically different – two estimates are statistically different at a given confidence level if

their confidence bounds do not overlap.

Overlay patterns can also be used to distinguish between statistically different estimates.

You can calculate whether confidence intervals overlap or include a fixed value using

selection queries.

Pros

-Identifies estimates that are statistically significant from another estimate, fixed value or CI.

Useful to compare to a state average or other important threshold.

Cons

-With small geographies, any overlay can obscure the underlying variable
-Some color/fill choices can be overpowering or difficult to read

Disclaimer?

Sample disclaimer (adapted from CDC):

One should not assume that the geographies mapped in different colors have significantly different rates. Estimates are grouped in categories by various methods to produce a map. The grouping does not incorporate standard deviation or confidence intervals and does not imply any formal comparisons between geographic units.

If you are concerned that map readers my make assumptions about significant differences between classes based on your map and you choose not to show reliability/uncertainty measures a disclaimer may be warranted