Uncertainty

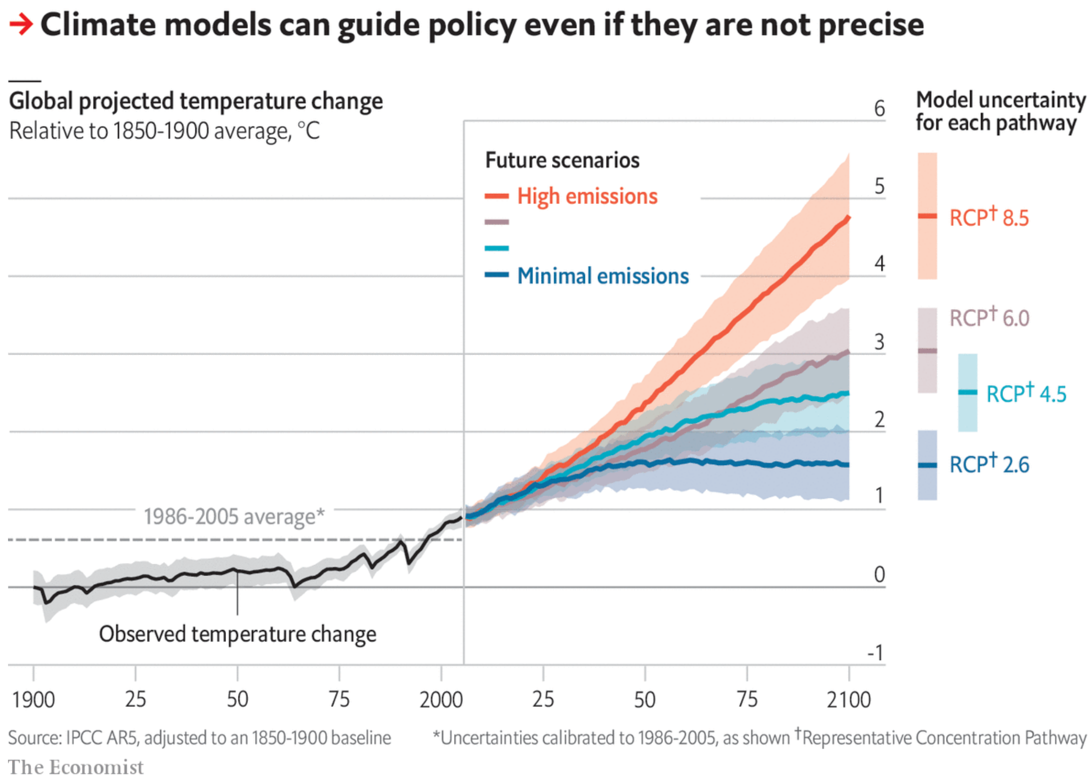
GEOG 5201 – Spring 2022

Outline

- Basic elements of uncertainty
  - Sources
    - Uncertainty in the raw data
    - Uncertainty in processing data
    - Uncertainty in the visualization
  - Concepts
    - Uncertainty versus error
    - Uncertainty, reliability, and quality
  - Why uncertainty matters in map making
    - Ethical necessity
    - Decision-making

Uncertainty in Maps

- We often think of maps as truthful representations of reality
  - This may not be correct, because the *truth* is often unknown to us, and the data visualized is very likely to deviate from the truth



Sources of Uncertainty

- Uncertainty is commonly used in the literature to describe the potential variation in values of an attribute at a spatial location
- Uncertainty emerges during multiple stages of map-making
  - Collection uncertainty -- uncertainty in the raw data
  - Derived uncertainty -- uncertainty in processing data
  - Visualization uncertainty -- uncertainty in the visualization

Sources of Uncertainty: Uncertainty in the Raw Data

- Example: sampling errors in the American Community Survey (ACS)
  - Because the ACS is based on a sample, rather than all housing units and people, ACS estimates have a degree of uncertainty associated with them, known as sampling error
  - The U.S. Census Bureau provides a margin of error (MOE) for each published ACS estimate to help understand the uncertainty

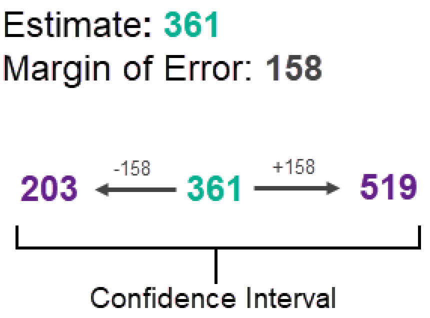
Table 7.1. **Sample Estimates and Margins of Error in American FactFinder: 2015**

Subject	Colorado			
	Estimate	Margin of Error	Percent	Percent Margin of Error
HOUSEHOLDS BY TYPE				
Total households	2,074,735	+/-7,548	2,074,735	(X)
Family households (families)	1,331,861	+/-11,075	64.2%	+/-0.5
With own children of the householder under 18 years	597,501	+/-8,075	28.8%	+/-0.4
Married-couple family	1,038,040	+/-9,389	50.0%	+/-0.4
With own children of the householder under 18 years	435,028	+/-7,158	21.0%	+/-0.3
Male householder, no wife present, family	93,024	+/-5,026	4.5%	+/-0.2
With own children of the householder under 18 years	48,969	+/-4,185	2.4%	+/-0.2
Female householder, no husband present, family	200,797	+/-6,582	9.7%	+/-0.3
With own children of the householder under 18 years	113,504	+/-4,699	5.5%	+/-0.2
Nonfamily households	742,874	+/-10,127	35.8%	+/-0.5
Householder living alone	564,757	+/-10,127	27.2%	+/-0.5
65 years and over	182,959	+/-5,249	8.8%	+/-0.3
Households with one or more people under 18 years	657,324	+/-8,777	31.7%	+/-0.4
Households with one or more people 65 years and over	497,903	+/-4,124	24.0%	+/-0.2

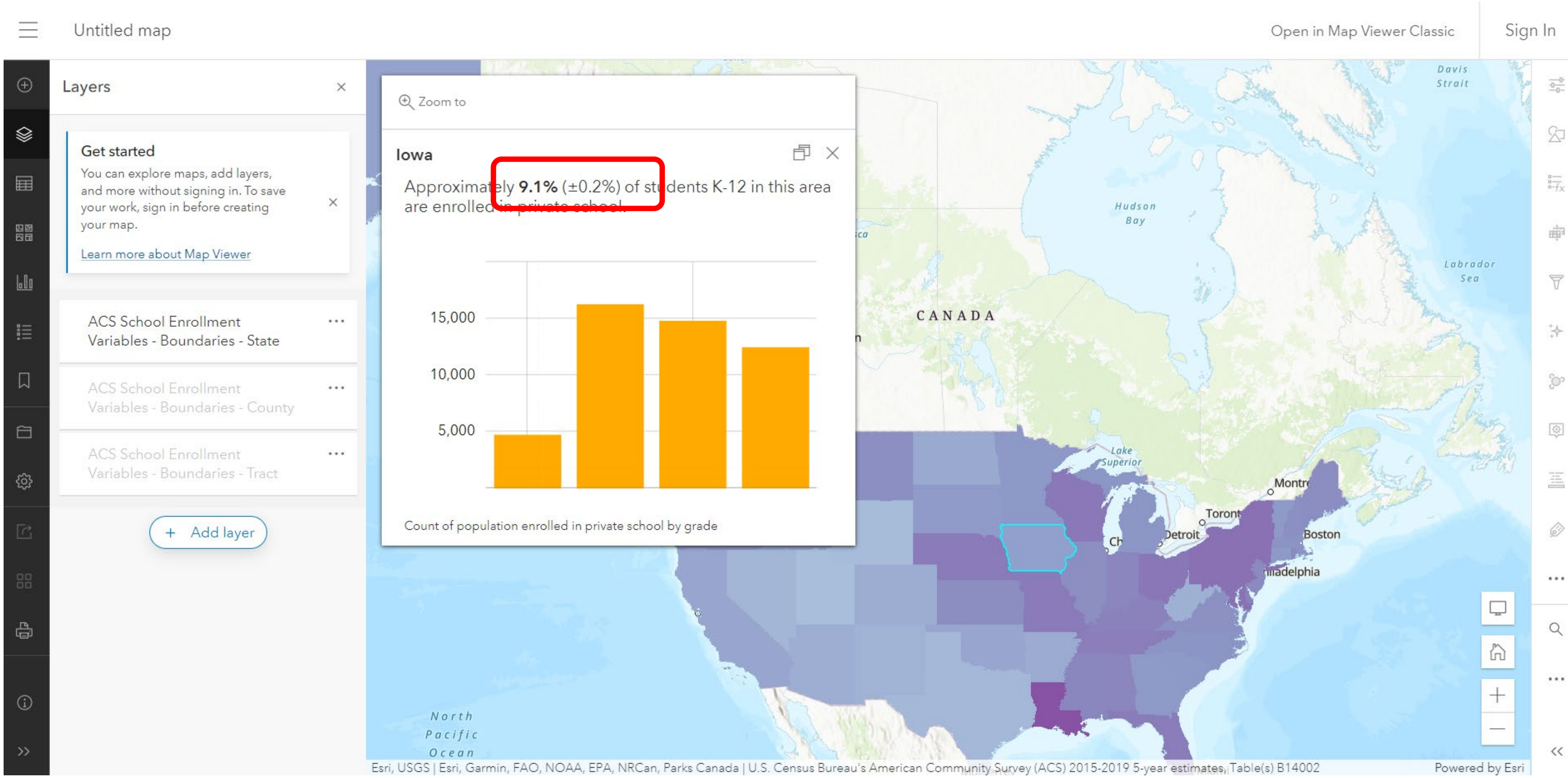
Source: U.S. Census Bureau, American FactFinder, Table DP02: Selected Social Characteristics in the United States.

Sources of Uncertainty: Uncertainty in the Raw Data

- How to interpret MOEs?
  - At a given confidence level, the MOE, combined with the ACS estimate, give users a range of values within which the actual, “real-world” value is likely to fall
    - 90% confidence level is the Census Bureau Standard
    - This range is called a confidence interval
  - Example: percentage of family households is 64.2% ± 0.5%
    - With 90% confidence, the range 63.7% to 64.7% covers the actual percentage of family households
    - This means that if the survey is conducted 100 times, 90 times the percentage of family households would be within 63.7% and 64.7%, and 10 times the percentage of family households would be either higher than 64.7% or lower than 63.7%



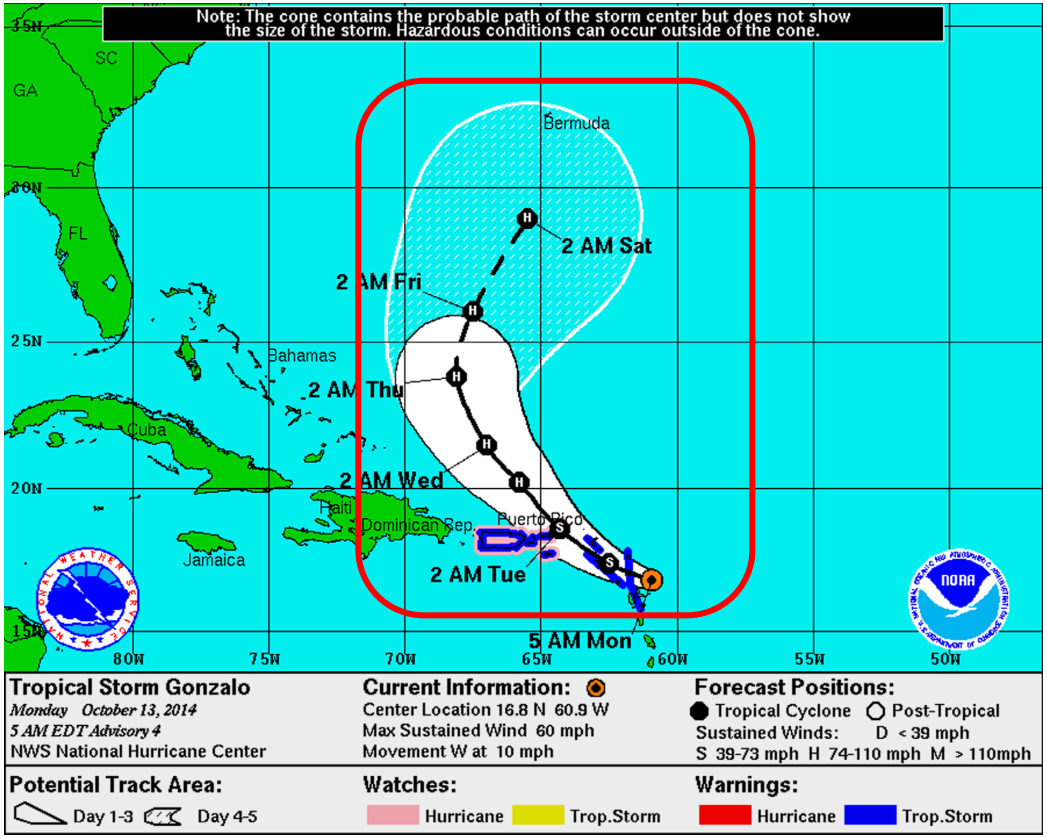
More on ACS uncertainty: [ACS data in ArcGIS Living Atlas]



A map showing both estimates of public vs. private school enrollment and associated MOEs

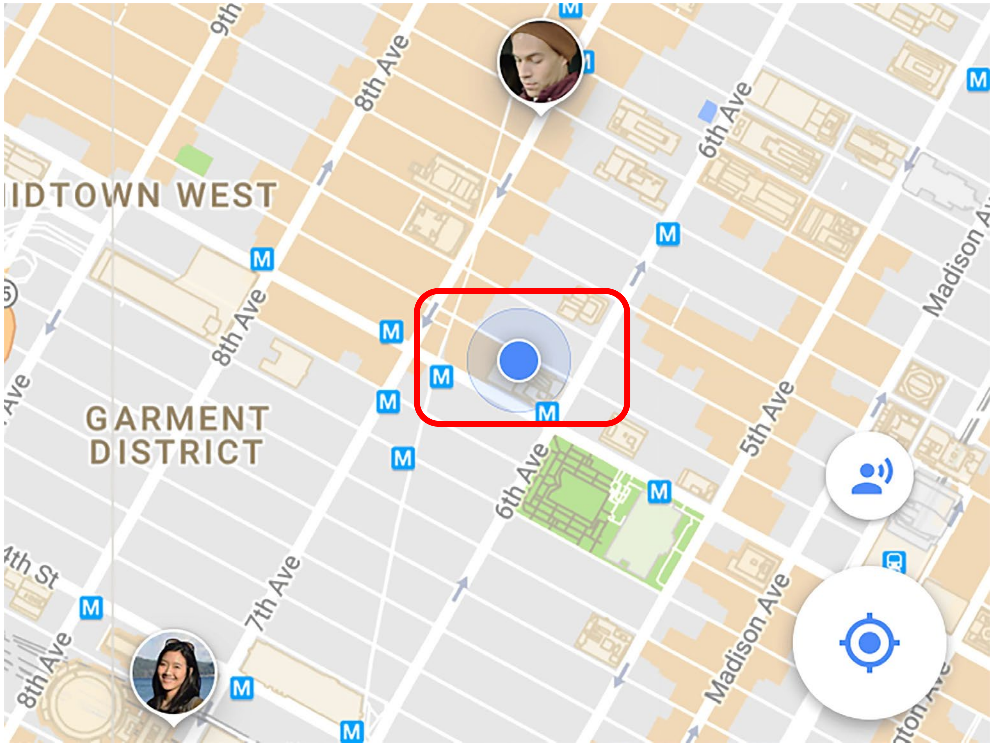
Sources of Uncertainty: Uncertainty in the Raw Data

- Example: “cone of uncertainty” in hurricane projections
  - The cone contains the probable path of the storm center



Sources of Uncertainty: Uncertainty in the Raw Data

- Example: circle of uncertainty on Google Maps
  - Google's documentation states that ". . .at times, you may see the dot surrounded by a light blue circle. This indicates that there is some uncertainty about your location." and ". . .you may be anywhere within it"

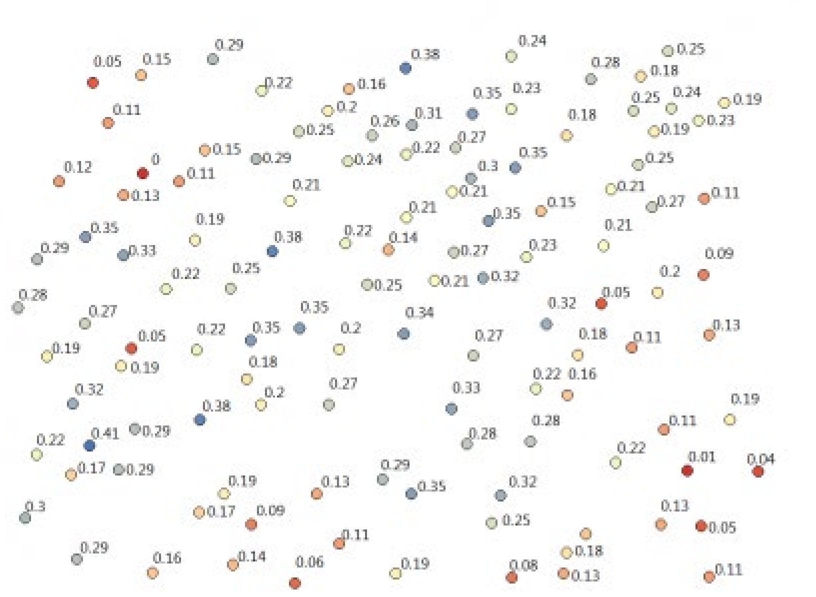


Question 3-1-1

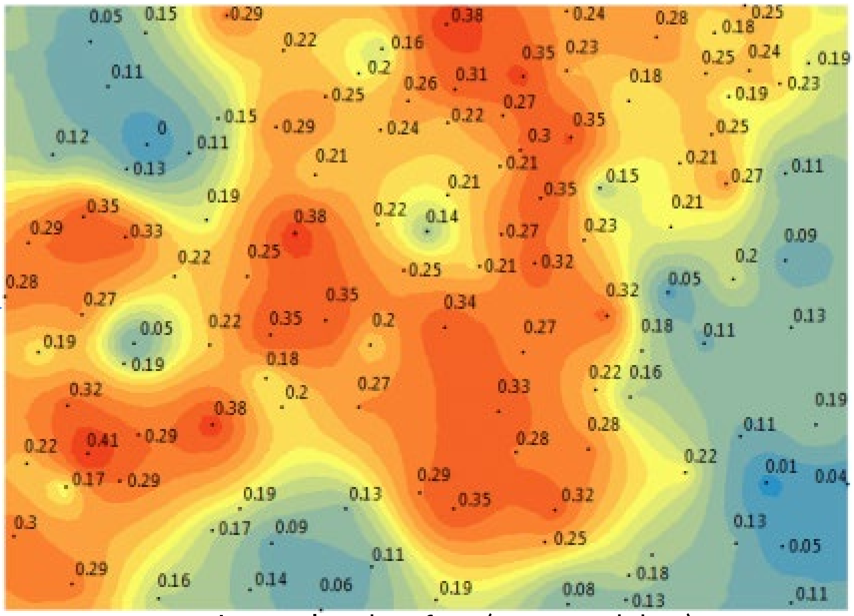
Recall any map you have previously seen or created, either in this class or elsewhere. Explain how uncertainty in the raw data can arise.

Sources of Uncertainty: Uncertainty in Processing Data

- Example: spatial interpolation
  - Use known point values to estimate unknown point values



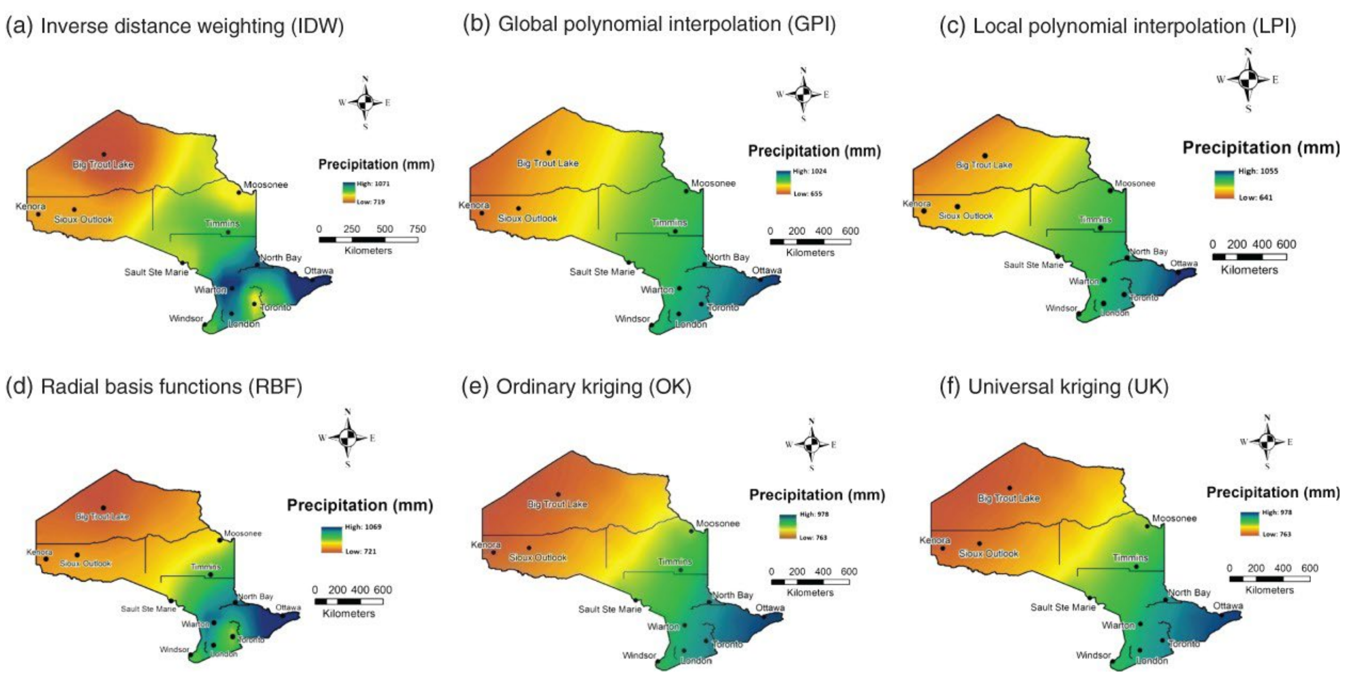
Known point values (raw data)



Interpolated surface (processed data)

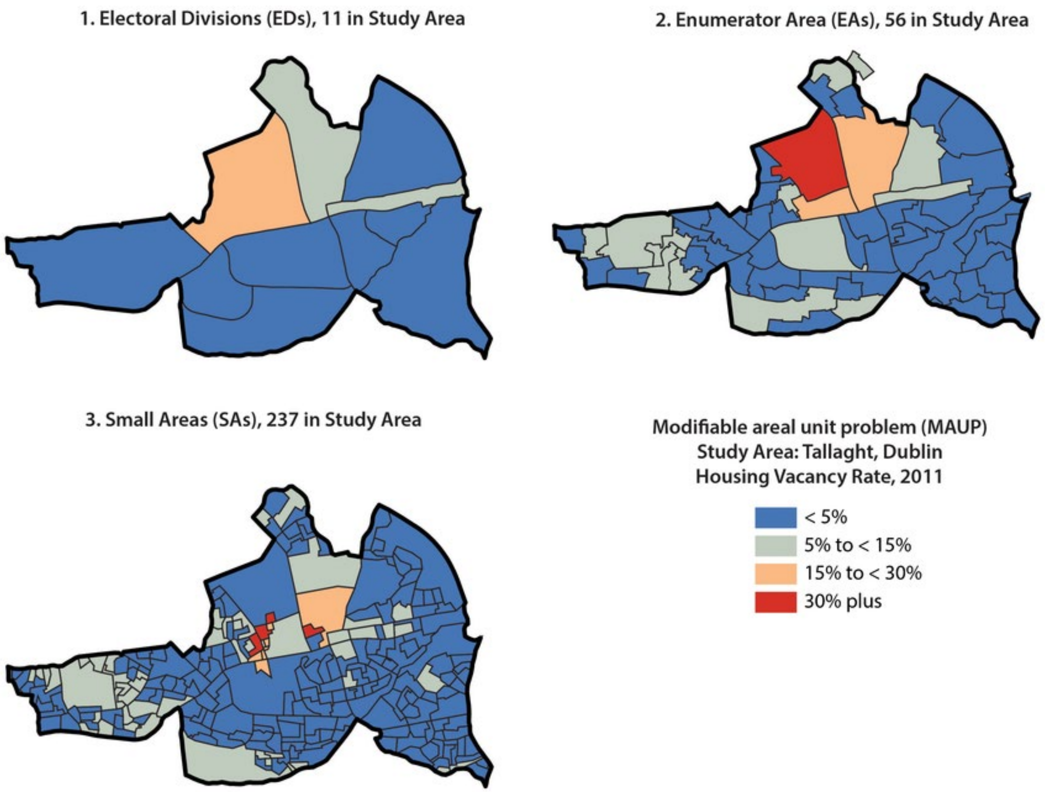
Sources of Uncertainty: Uncertainty in Processing Data

- There are a variety of algorithms, each producing a potentially different set of interpolated values
- For any particular location on a map, we can consider the set of interpolated values to be the data's uncertainty



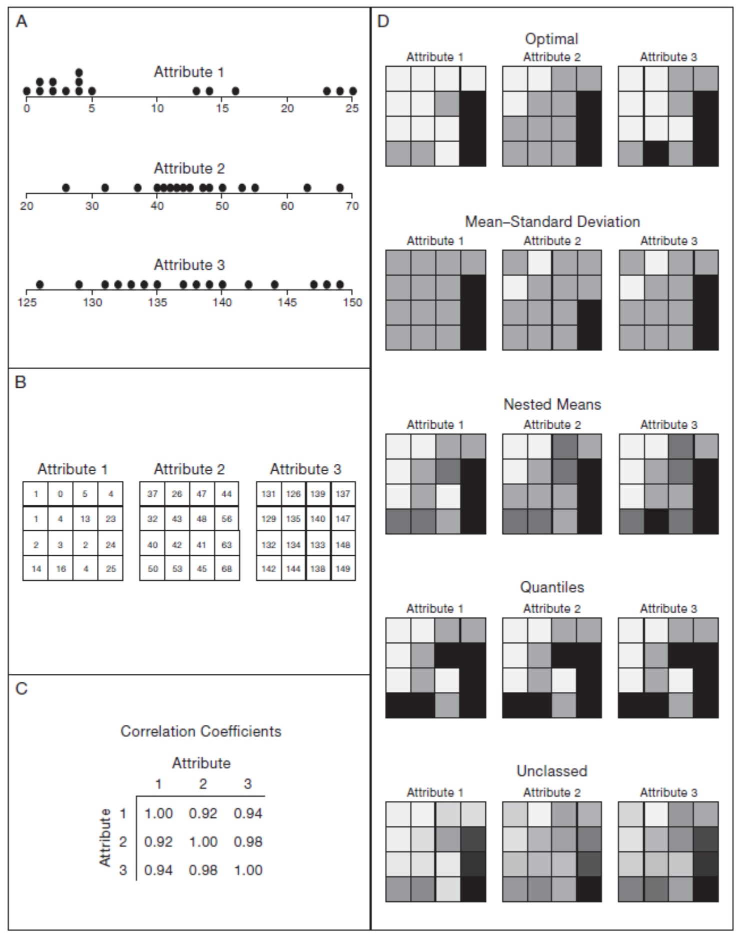
Sources of Uncertainty: Uncertainty in Processing Data

- Example: the modifiable areal unit problem (MAUP)
  - The same basic data yield different results when aggregated in different ways



Sources of Uncertainty: Uncertainty in the Visualization

- Example: different classifications lead to different visualization results



Concepts of Uncertainty: Uncertainty versus Error

- Recall that uncertainty is the potential variation in values of an attribute at a spatial location
  - True values typically unknown
- Error is the difference between the measured value and the true value of an attribute at a spatial location
  - True values are known objectively
- Uncertainty covers a broader range of doubt than error alone

Concepts of Uncertainty: Uncertainty, Reliability, and Quality

- The terms “reliability” and “quality” are also used
  - *Uncertainty* is equated with *unreliability* and *poor quality*
- The U.S. Federal Information Processing Standard 173 lists 5 categories for assessing data quality
  - Lineage: history of data, including sources, data processing and transformations
  - Position accuracy: location accuracy of geographic features (recall the hurricane projection and Google Maps examples)
  - Attribute accuracy: accuracy of features found at particular locations (recall the ACS example)
  - Logical consistency: extent to which objects within the dataset agree; topological correctness
  - Completeness: extent to which data is comprehensive

Question 3-1-2

Explain the difference between uncertainty and error using any of the examples above (p.5 – p.14).

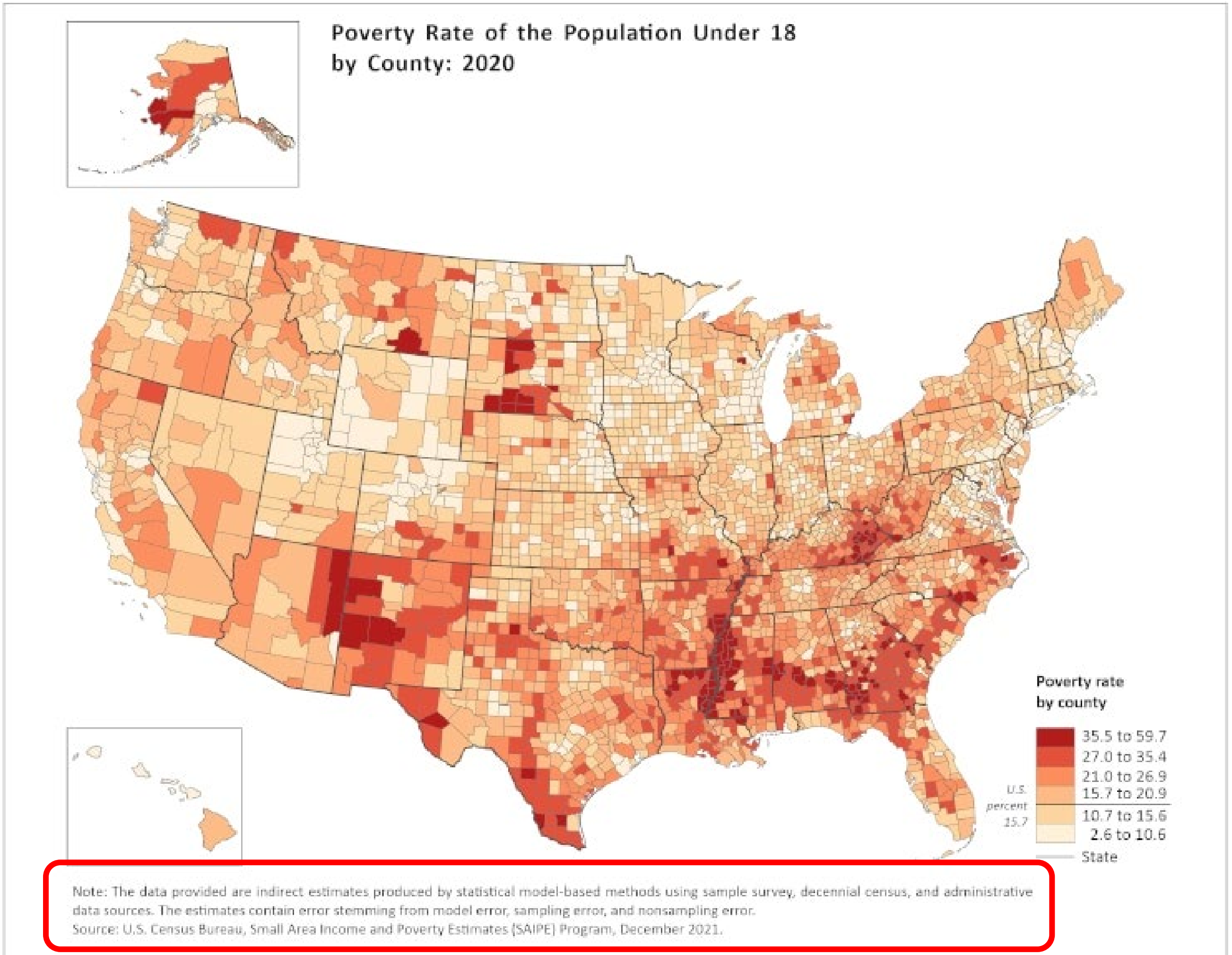
Why Uncertainty Matters in Map Making: Ethical Necessity

- To indicate the (unavoidable) gap between reality and representation
  - Users need to understand that GIS data and analyses are not necessarily accurate and reliable
  - GIScience researchers, educators and students should know that there is a multitude of reasons that our representations are incomplete and uncertain
- Withholding the uncertainty information from map readers would be misleading
- Good science includes statements of accuracy, and the reliability of results must be understood and communicated


Mad scientists are cloning dinosaurs as weapons of the future

Monsters would be last line of defense in nuclear war

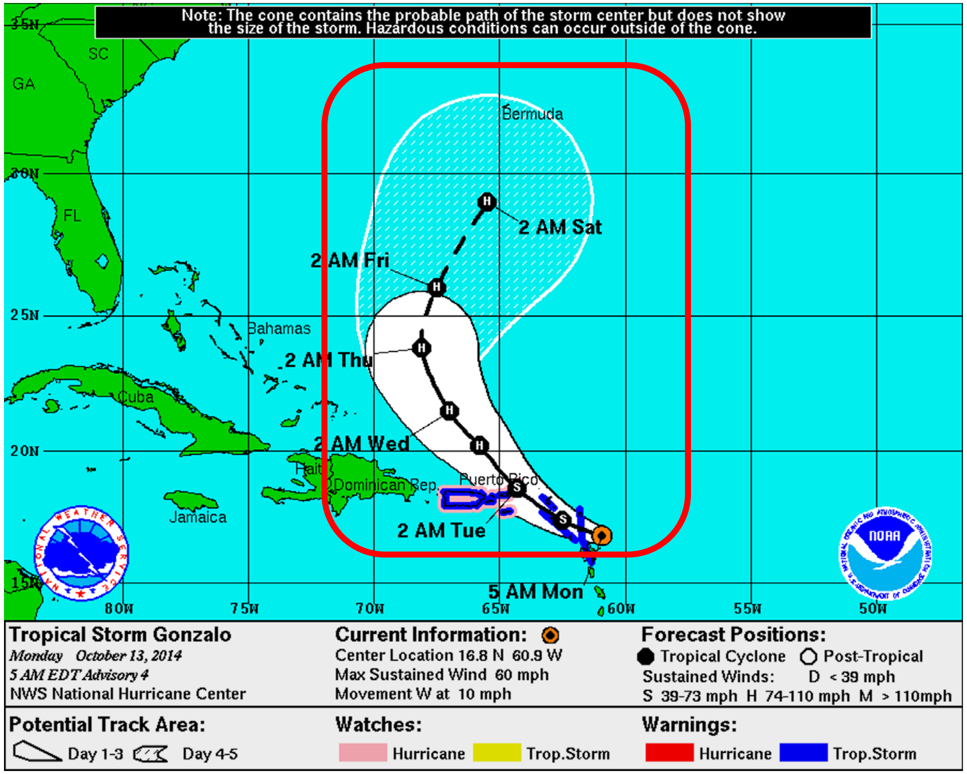
Map of poverty rate of the population under 18 by the U.S. Census Bureau *without uncertainty information presented*.



The data provided are indirect estimates by statistical model-based methods using sample survey, but users may interpret them as population enumeration results

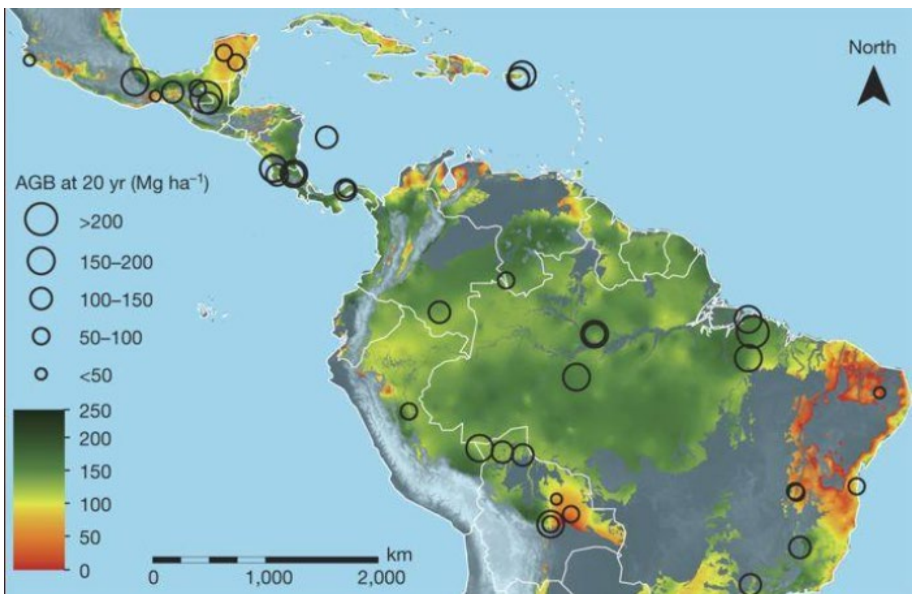
Why Uncertainty Matters in Map Making: Decision-Making

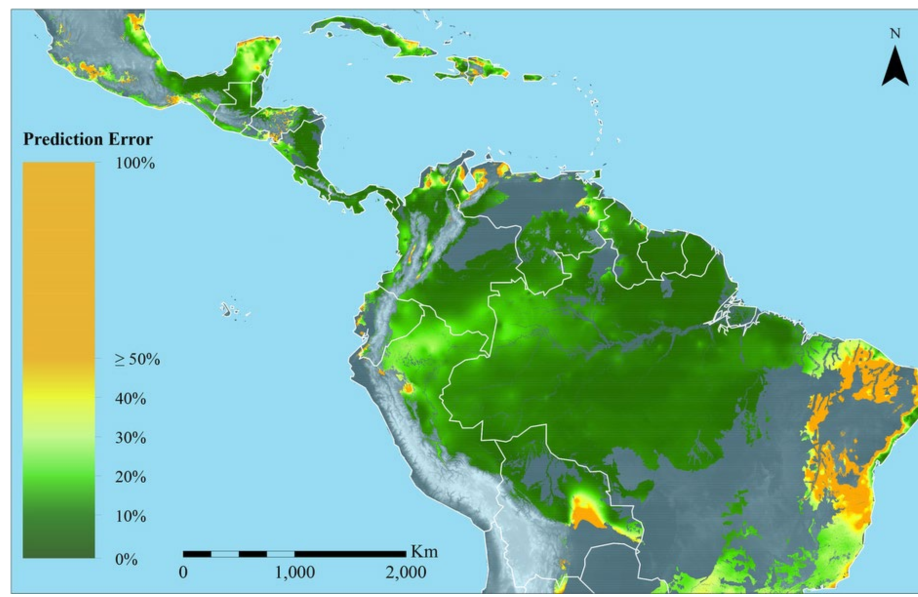
- Uncertainty plays an important role in decision-making
  - Example: the “cone of uncertainty” in maps of predictive hurricane paths often play an important role in decisions made by residents of storm-affected areas



Why Uncertainty Matters in Map Making: Decision-Making

- Professional fields dealing with the natural environment routinely require their practitioners to make decisions using data that can include a range of uncertainties
  - Example: Uncertainty map of potential biomass recovery of Neotropical secondary forests





Question 3-1-3

Can you think of a map where the uncertainty information could be useful? Why do you think uncertainty matters in that map?