VIS 2019

Exploring the Sensitivity of Choropleths under Attribute Uncertainty

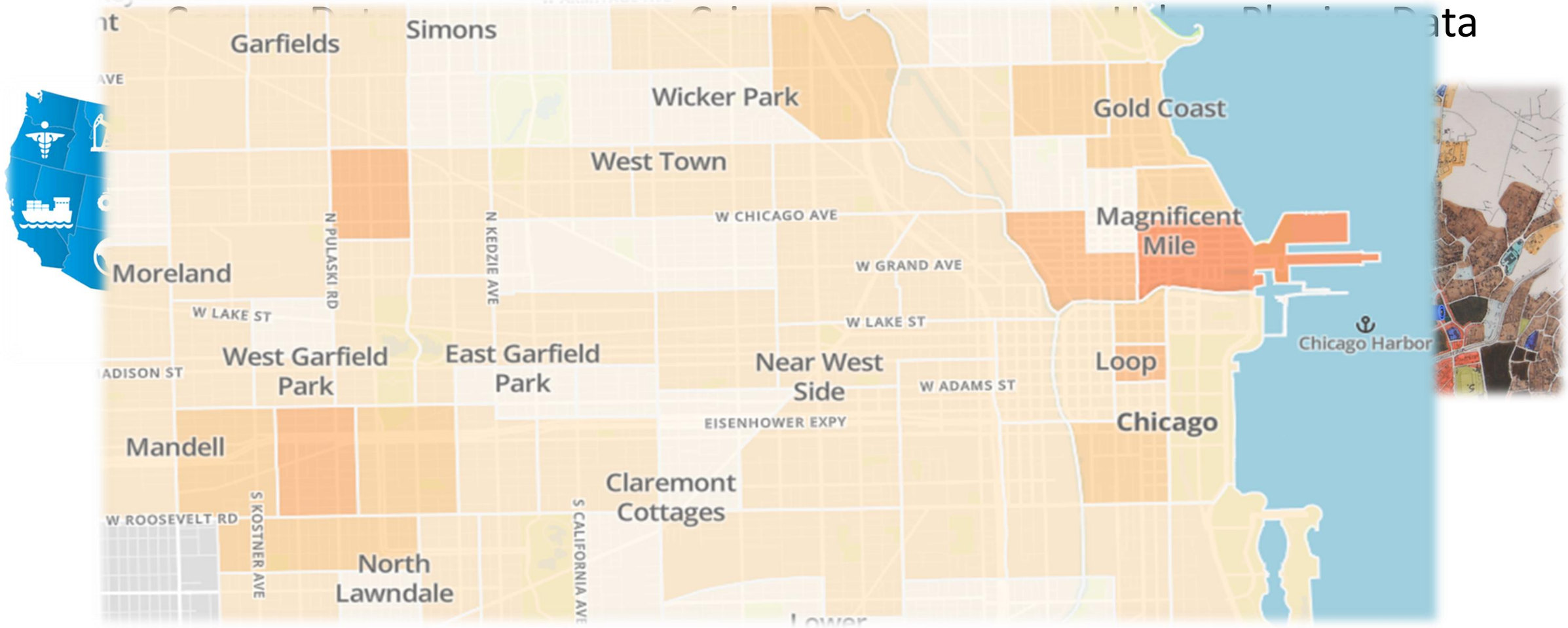
Zhaosong Huang[1], Yafeng Lu[2], Elizabeth A. Mack[3],
Wei Chen[1], and Ross Maciejewski[2]

1 The State Key Lab of CAD&CG, Zhejiang University
2 The School of Computing, Informatics & DSE, Arizona State University
3 The Department of Geography, Michigan State University





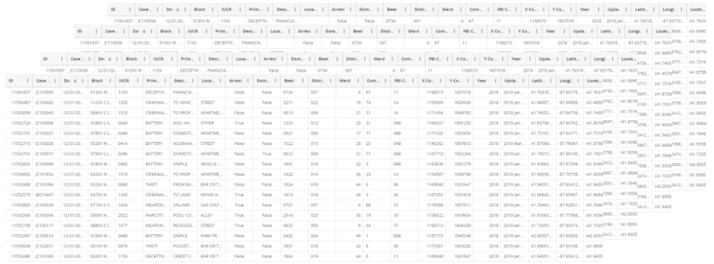
Geospatial Analysis

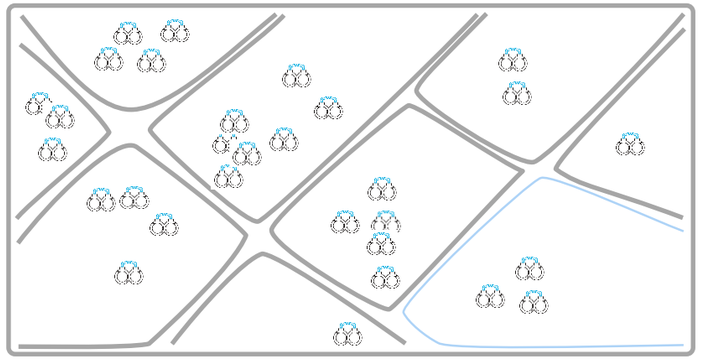


The key of Choropleth map

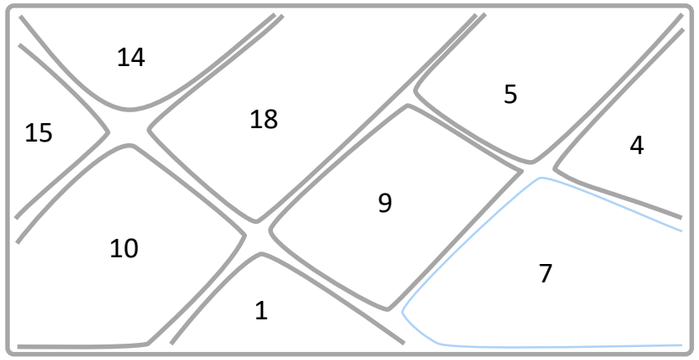
Choropleth map is one of the most common methods of visualizing spatially referenced data

Raw data

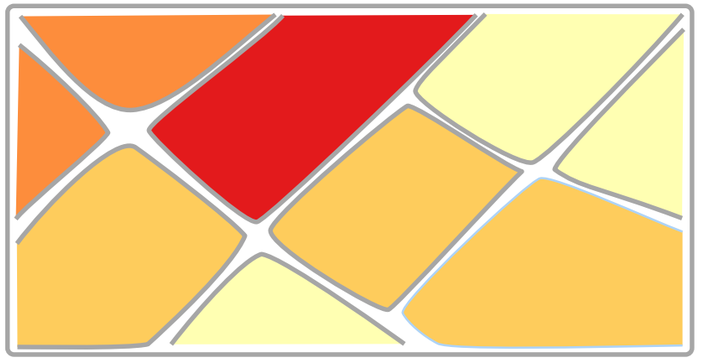


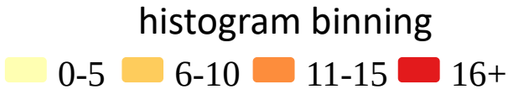


Statistical data



Choropleth Map

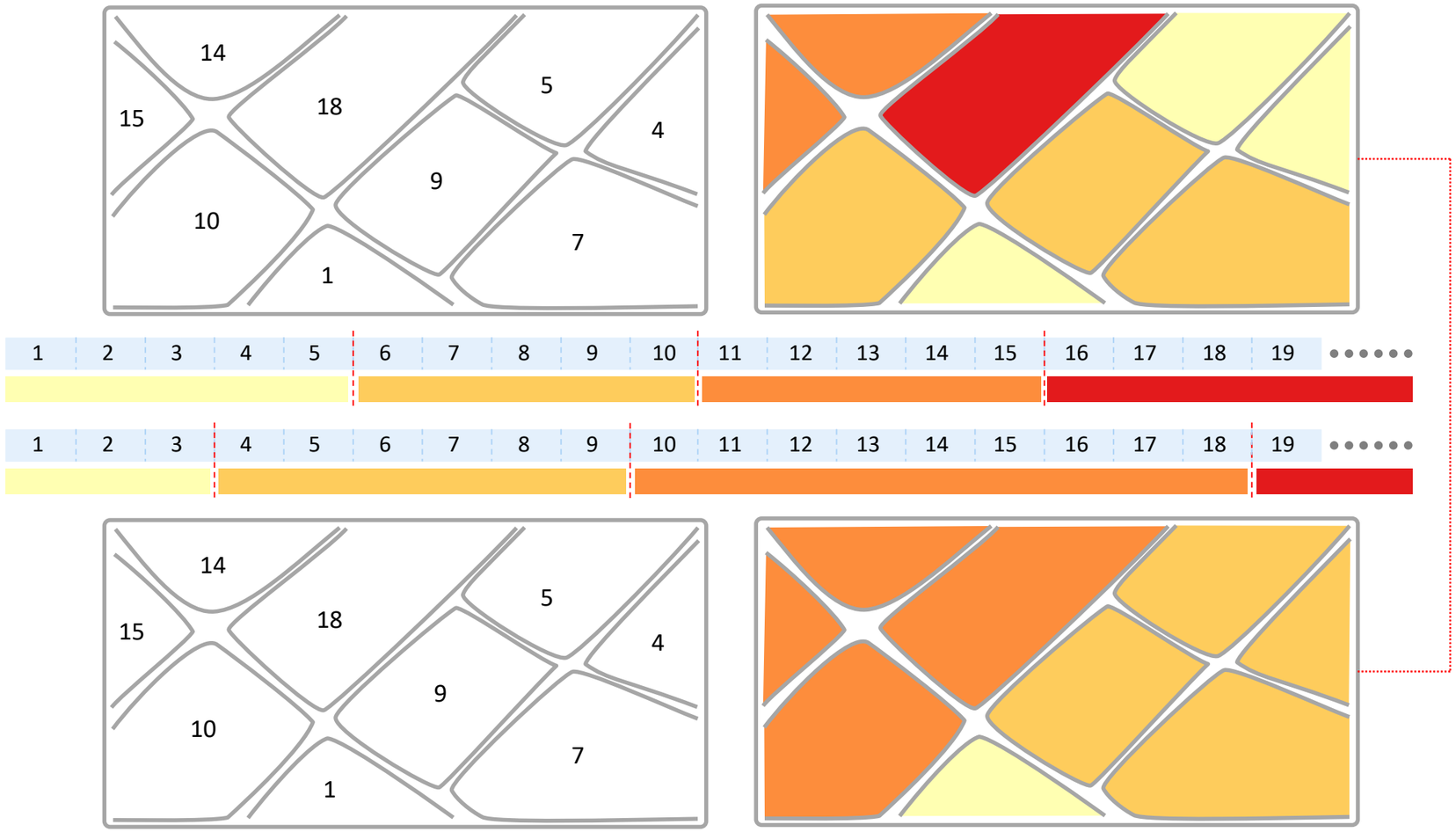




- Based on statistical data
- Regions are classified into several groups
- Region color is decided by a histogram binning of the measured variable

Choropleth Maps – What happens under uncertainty?

The visual representation of the choropleth map is highly influenced by the class interval selection

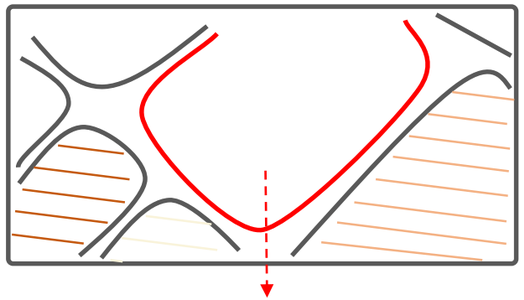


Choropleth Maps – Data Uncertainty

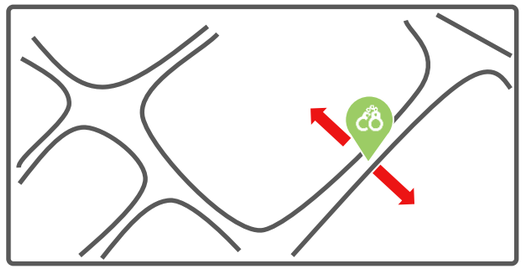
Many geographic datasets have an inherent level of uncertainty

GPS traces Twitter locations American Community Survey[1]

In choropleth maps, we are mostly interested in attribute uncertainty



Around 10 Crimes



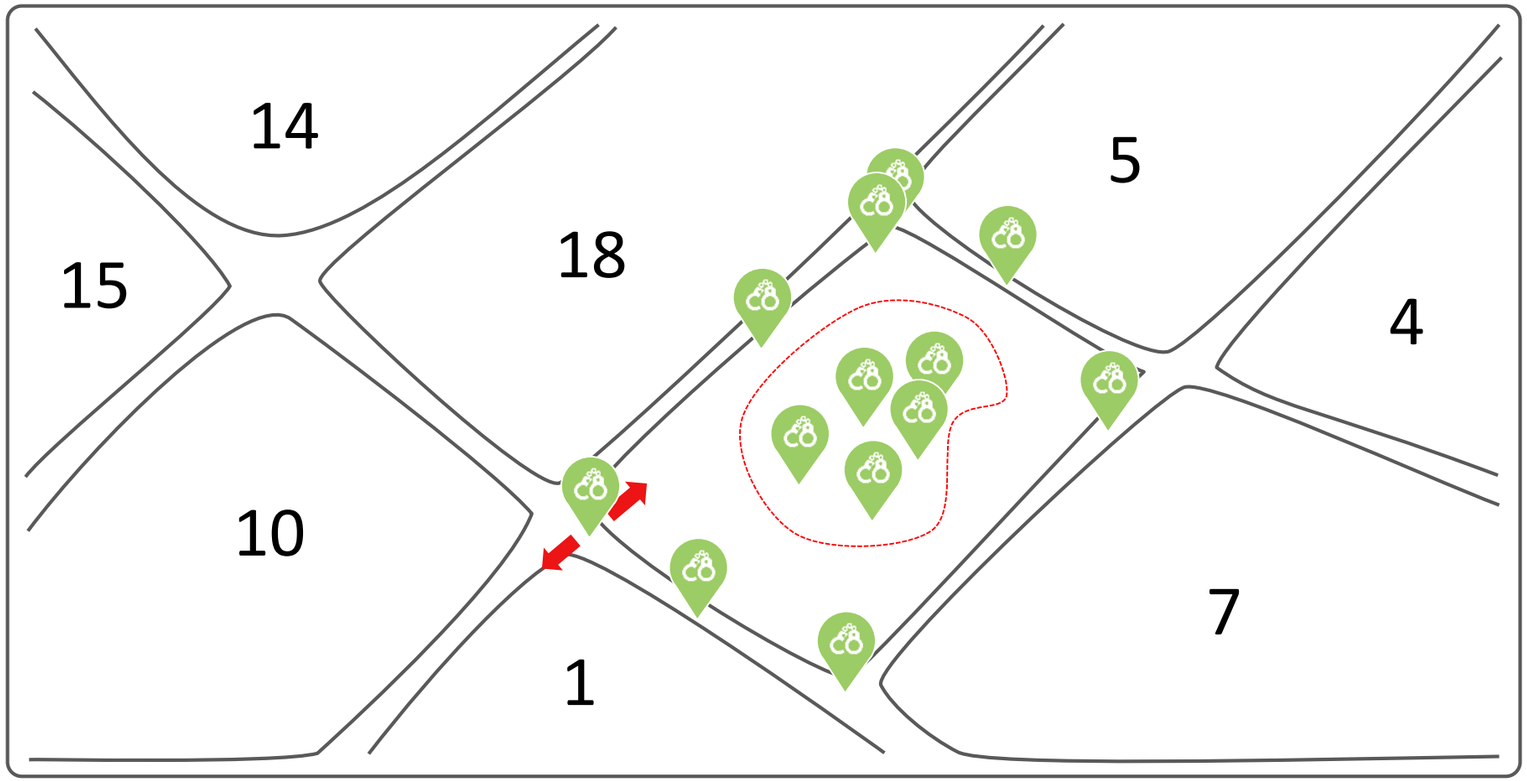
Ambiguous Location

- Arises from varying the spatial aggregation of data
- Often referred to as the modifiable areal unit problem (MAUP)[2]

1 - H. MacDonald. The American Community Survey: Warmer (more current), but fuzzier (less precise) than the decennial census. Journal of the American Planning Association, 72(4):491–503, 2006.
2 – C. E. Gehlke and K. Biehl. Certain effects of grouping upon the size of the correlation coefficient in census tract material. Journal of the American Statistical Association, 29(185A):169–170, 1934

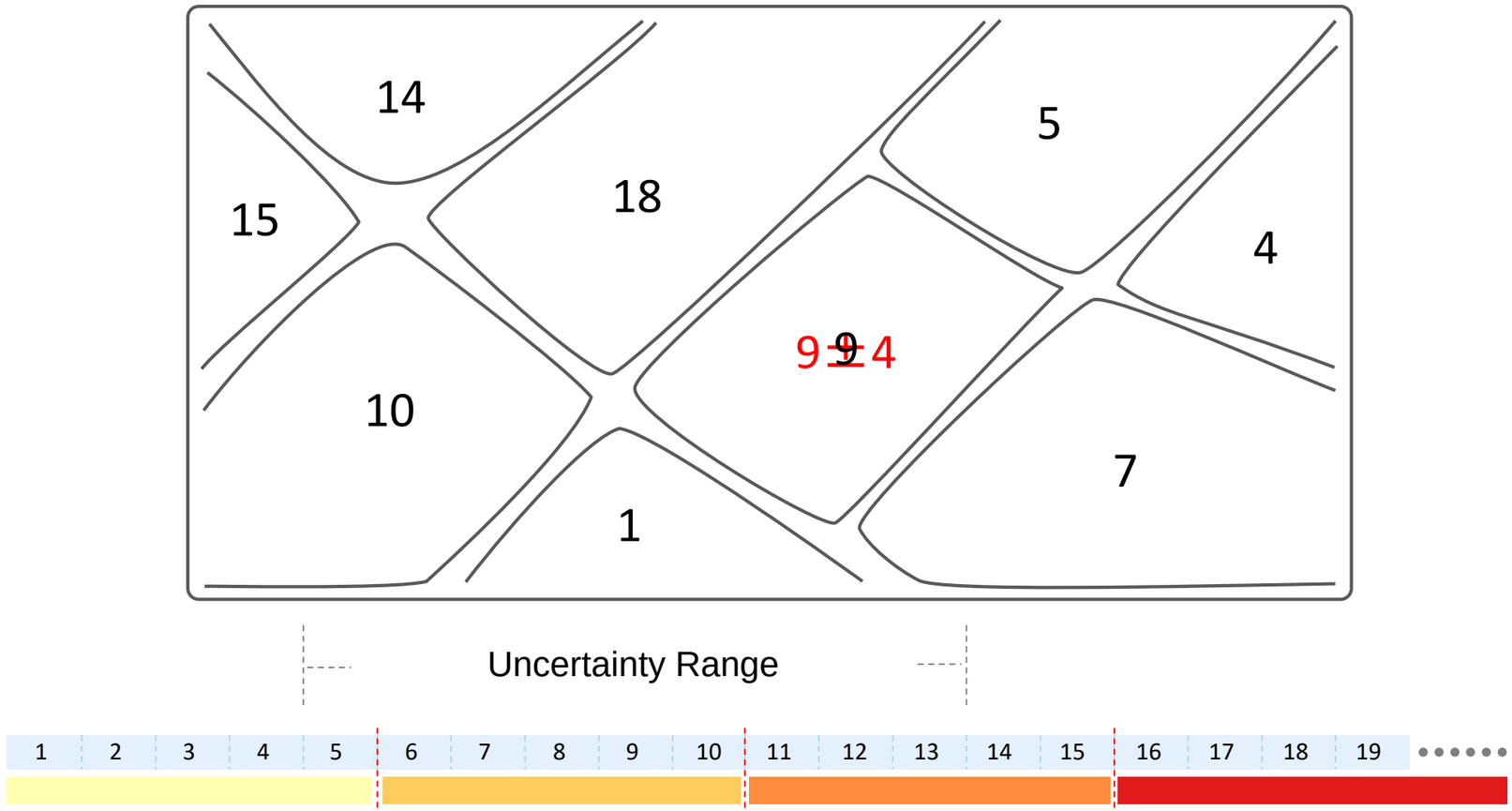
Choropleth Maps – What happens under uncertainty?

The visual representation of the choropleth map is highly influenced by the data uncertainty



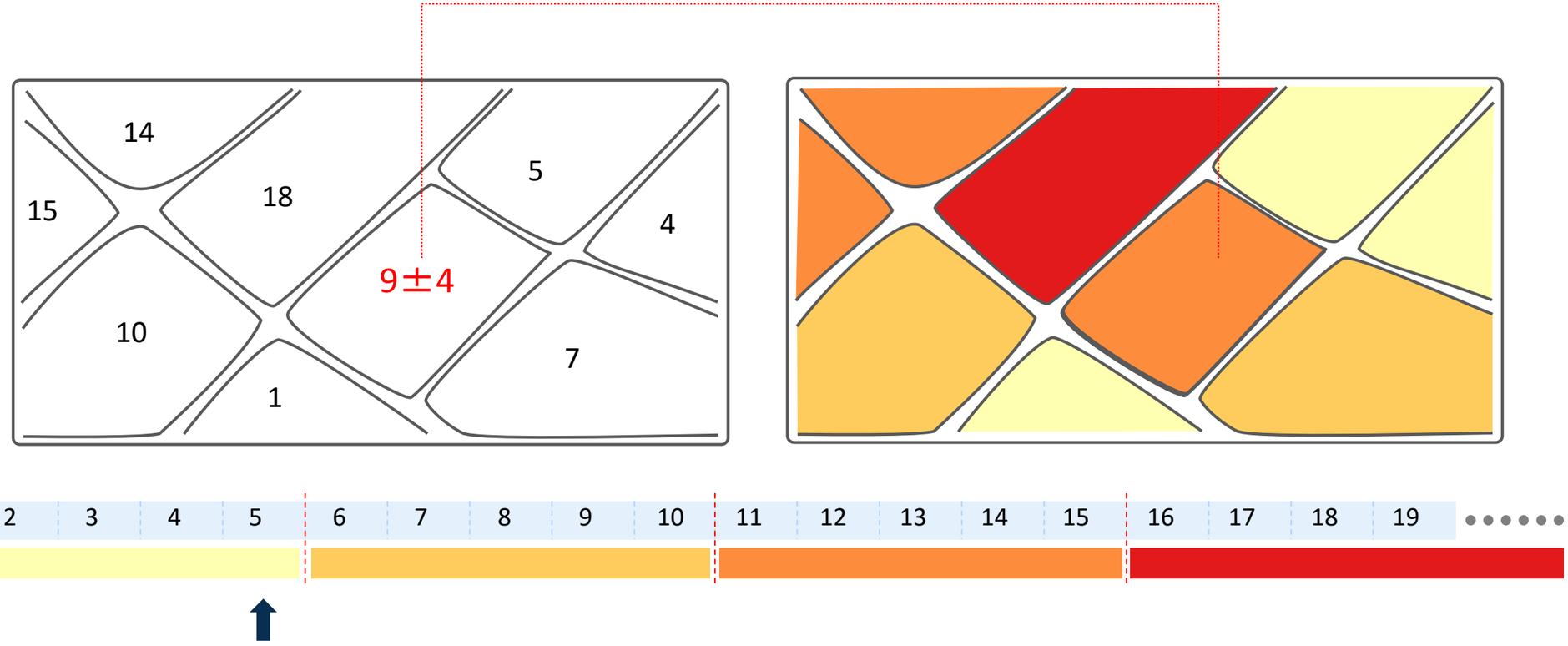
Choropleth Maps – What happens under uncertainty?

The visual representation of the choropleth map is highly influenced by the data uncertainty



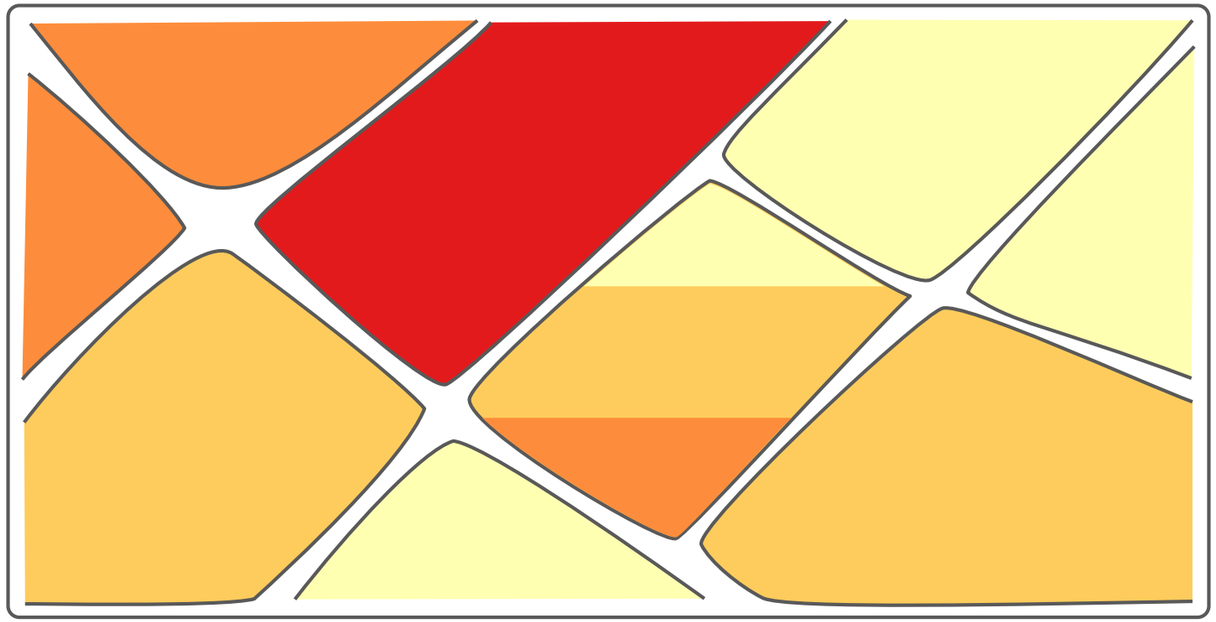
Choropleth Maps – What happens under uncertainty?

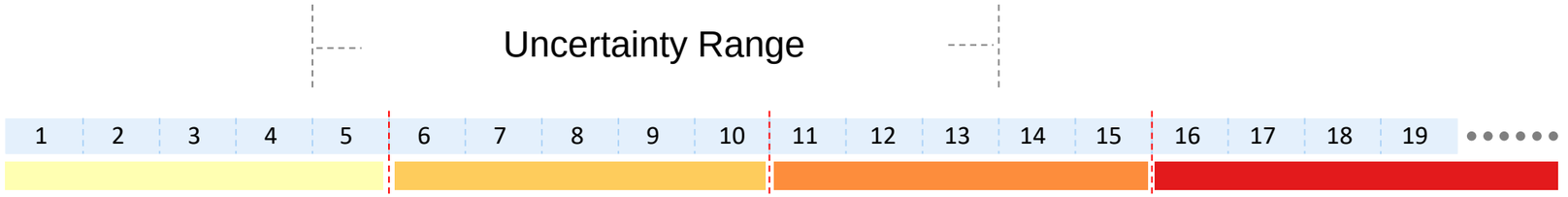
The visual representation of the choropleth map is highly influenced by the data uncertainty



Choropleth Maps – What happens under uncertainty?

What should the map look like?



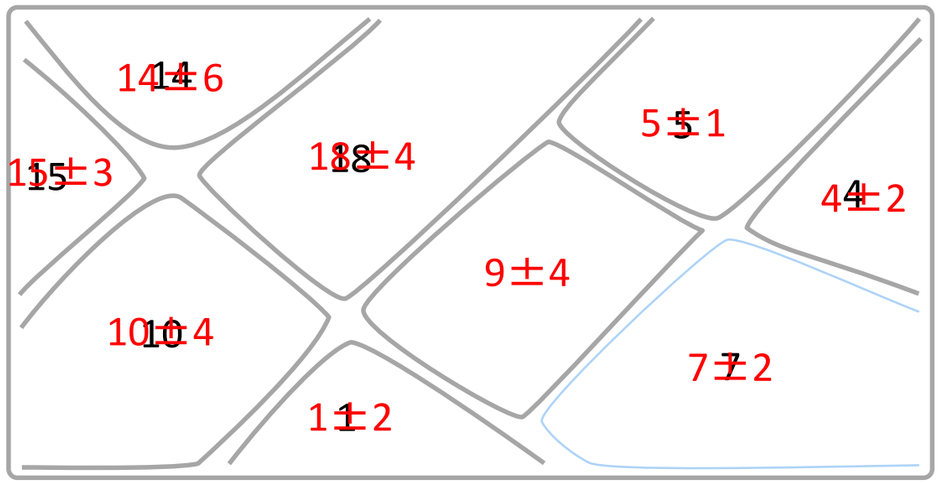


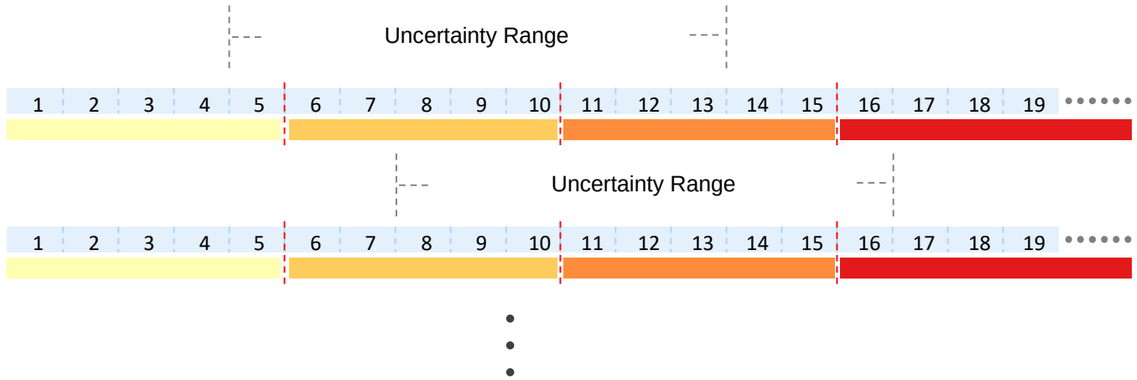
- Visual appearance
- Spatial autocorrelation
- Relationships between variables

Choropleth Maps – What happens if…?

The issue of uncertainty becomes even more complex in analyses implementing

Univariate context





Multivariate Classification

Variable name	Description
PST045214	Population, 2014 estimate
PST040210	Population, 2010 (April 1) estimates base
PST120214	Population, percent change - April 1, 2010 to July 1, 2014
POP010210	Population, 2010
AGE135214	Persons under 5 years, percent, 2014
AGE295214	Persons under 18 years, percent, 2014
AGE775214	Persons 65 years and over, percent, 2014
EDU635213	High school graduate or higher, percent of persons age 25+, 200
EDU685213	Bachelor's degree or higher, percent of persons age 25+, 2009-2
VET605213	Veterans, 2009-2013
LFE305213	Mean travel time to work (minutes), workers age 16+, 2009-2013
HSG010214	Housing units, 2014
HSG445213	Homeownership rate, 2009-2013
HSG096213	Housing units in multi-unit structures, percent, 2009-2013
HSG495213	Median value of owner-occupied housing units, 2009-2013
HSD410213	Households, 2009-2013

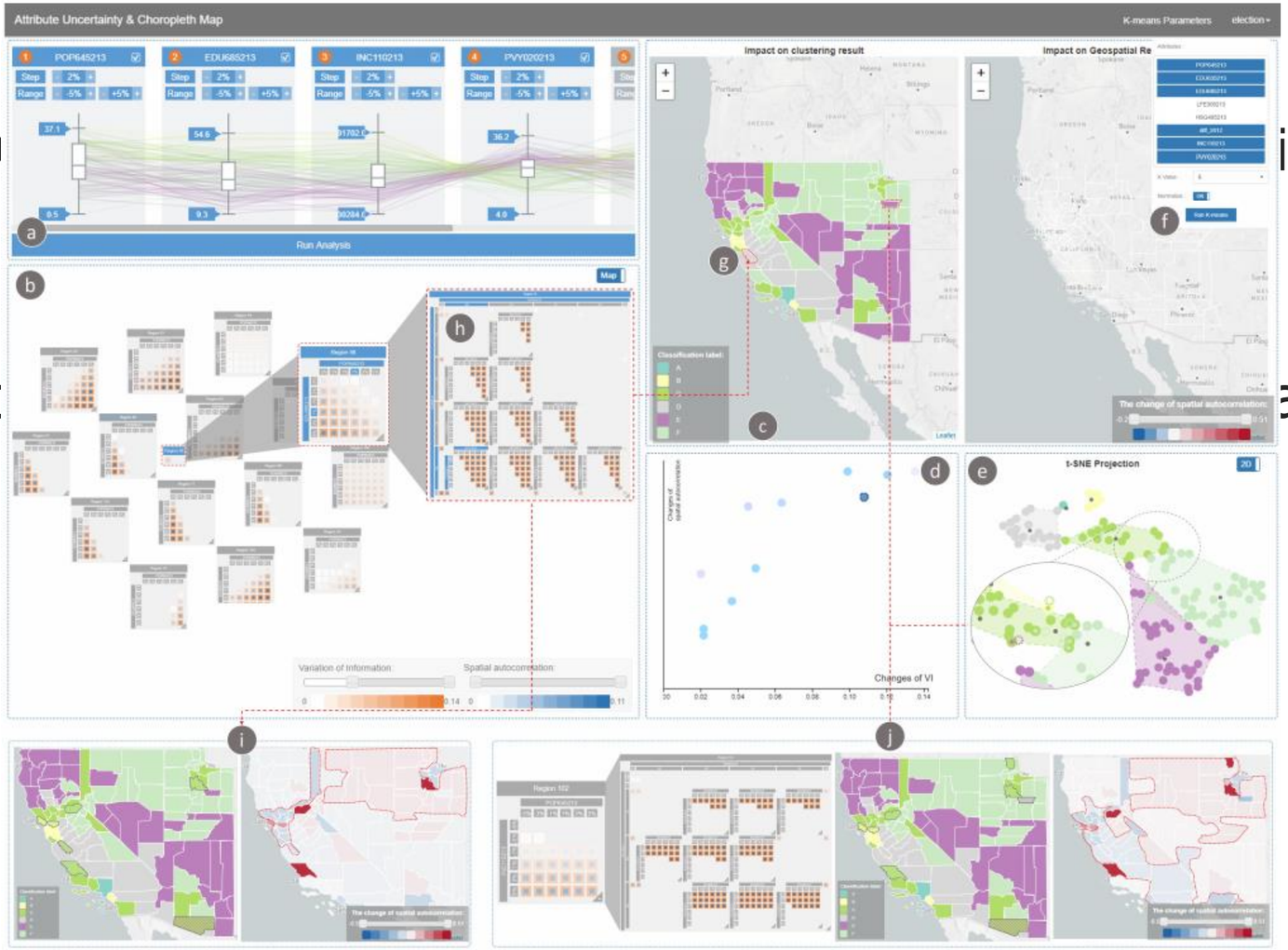
- It has been widely applied in various domains over various geographical areas
  - Regional ecosystems[2]
  - Demographic maps[1]

VIS 2019

1. Vickers, D. & Rees, P. Creating the UK National Statistics 2001 output area classification. *Journal of the Royal Statistical Society: Series A*. 2007.
2. Hargrove, W. W. & Hoffman, F. M. Potential of Multivariate Quantitative methods for Delineation and Visualization of Ecoregions. *Environmental Management*, 2005

Challenges

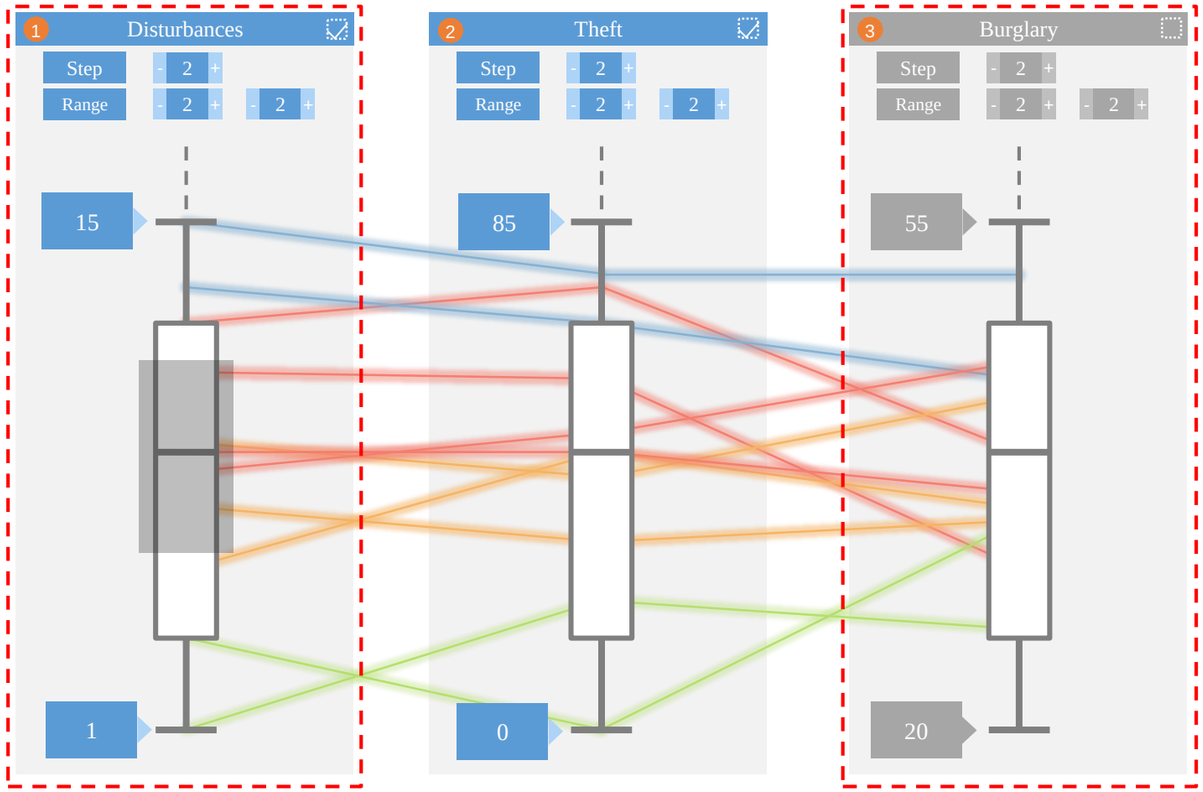
- D ign.
- Explo arance.



Specification of uncertainty of an attribute

Using a preferred clustering algorithm and Setting the uncertainty range

1. Show the distribution of the attribute's
2. Specify a range the value to change in
3. Specify the step the value to change in



Quantify the impact of data uncertainty

For each spatial unit, we simulate a classification across a range of uncertainty

- How many labels will change in the map if a measurement is uncertain
- How much spatial autocorrelation changes under the range of uncertainty

Quantify the impact of data uncertainty

- How many labels will change in the map if a measurement is uncertain
- How much spatial autocorrelation changes under the range of uncertainty

  - Join Count statistics[1]
  - Geary's C[2]
  - Getis-Ord General G[3]
  - Moran's I[4] is defined as

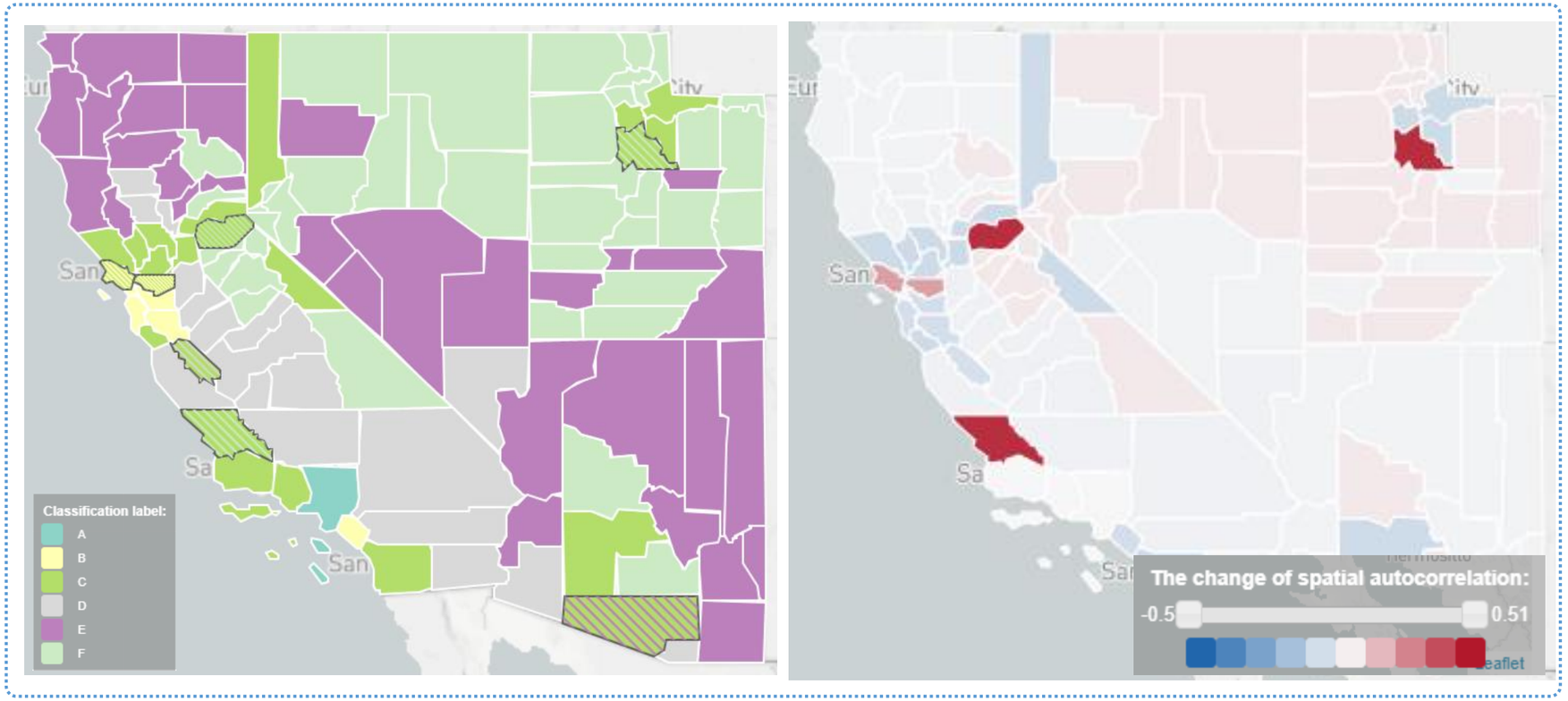
$$I = \frac{N}{\Sigma_i \Sigma_j w_{ij}} \frac{\Sigma_i \Sigma_j w_{ij}(x_i - \bar{X})(x_j - \bar{X})}{\Sigma_i (x_i - \bar{X})^2},$$

1. A. D. Cliff and J. K. Ord. *Spatial autocorrelation, vol. 5*. Pion London,1973
2. R. Geary. The contiguity ratio and statistical mapping. *The incorporated statistician*, 5(3):115–146, 1954.
3. A. Getis and J. K. Ord. The analysis of spatial association by use of distance statistics. *Geographical Analysis*, 24(3):189–206, 1992
4. P. A. Moran. Notes on continuous stochastic phenomena. *Biometrika*, pp.17–23, 1950

Quantify the impact of data uncertainty

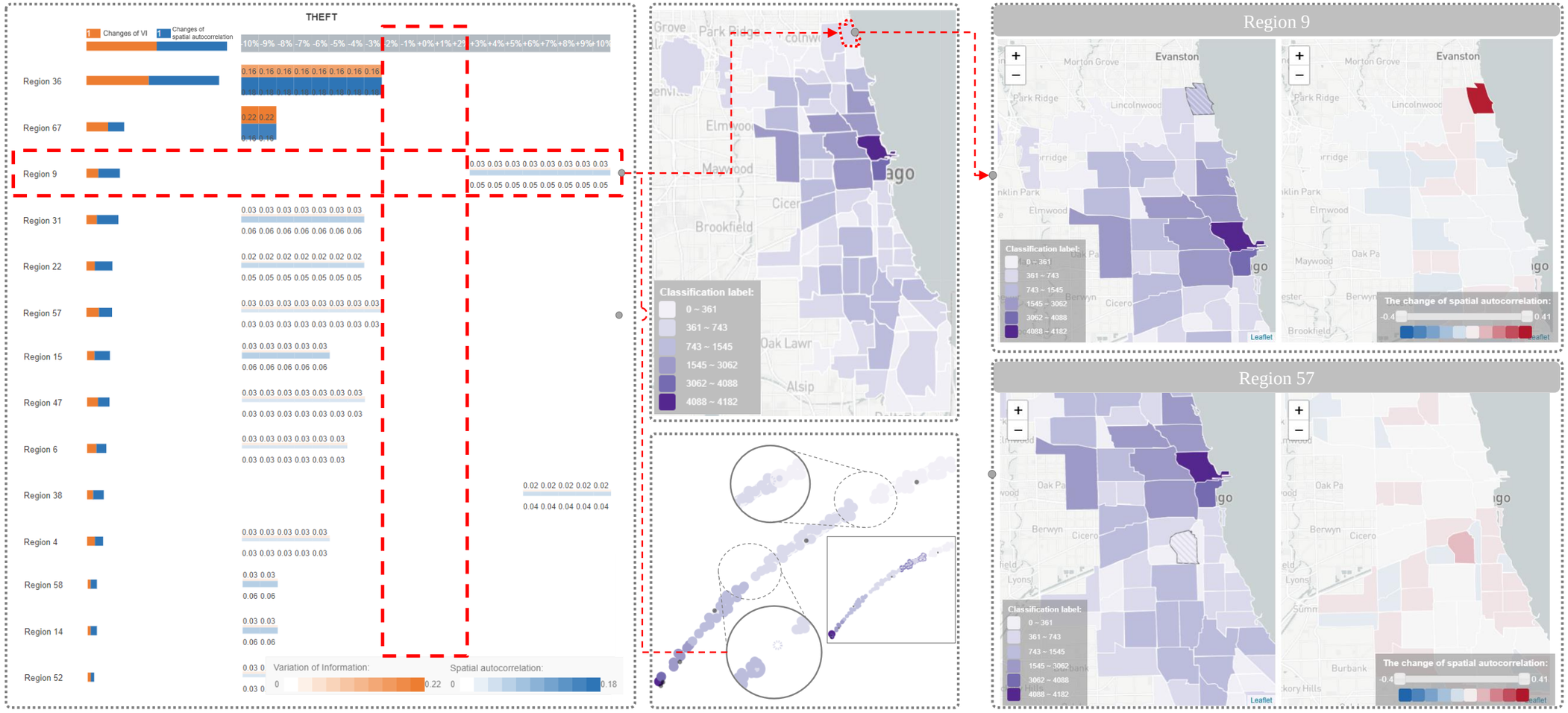
- How many labels will change in the map if a measurement is uncertain
- How much spatial autocorrelation changes under the range of uncertainty

Dual-Choropleth Map



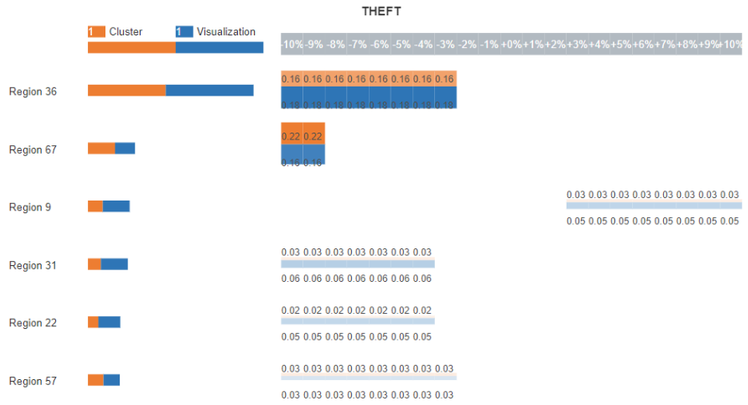
Exploring the Sensitivity of Choropleths under Attribute Uncertainty

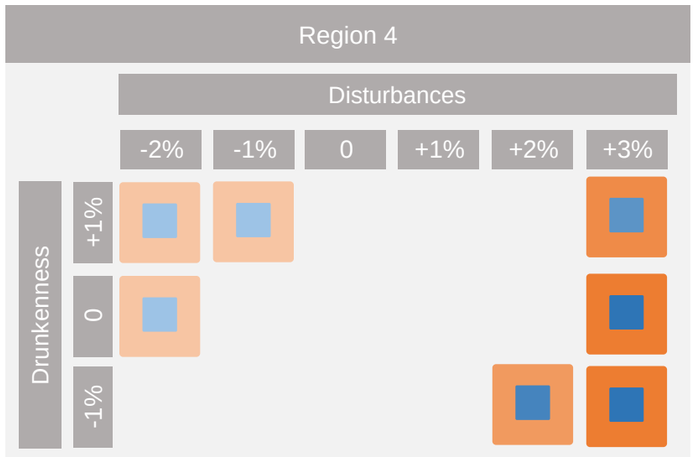
Single Attribute Impact Profile -- Impact River

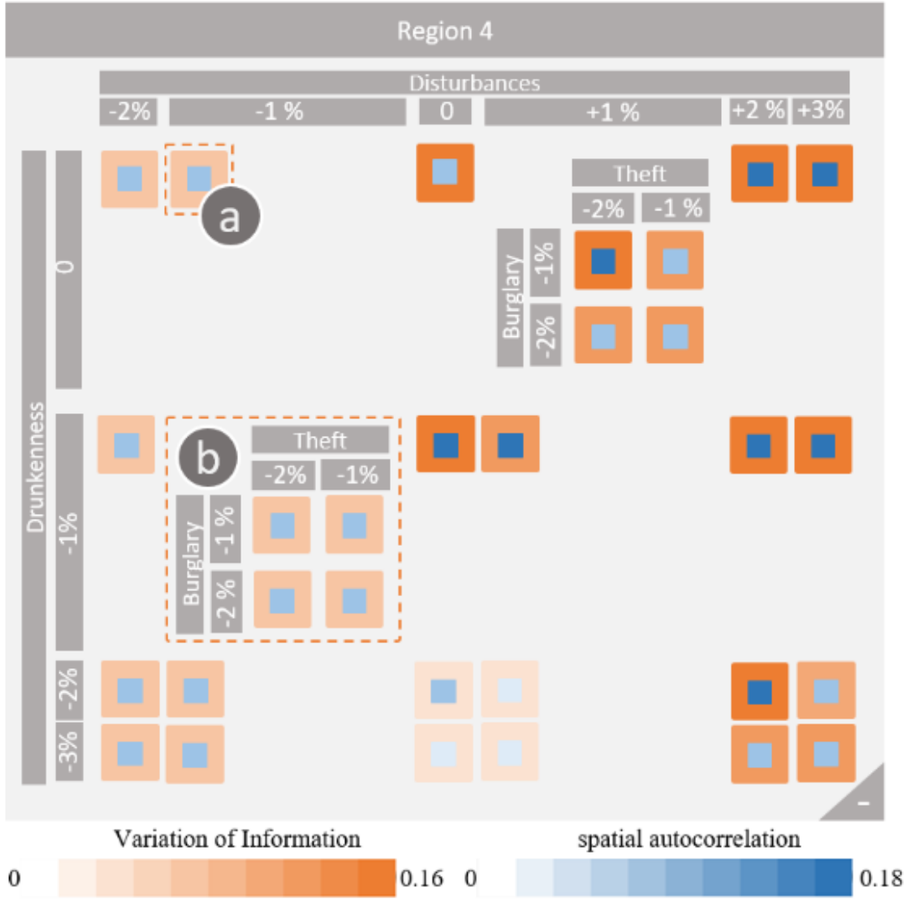


Exploring the Sensitivity of Choropleths under Attribute Uncertainty

Impact Matrix – Multi-Attribute Impact Profile



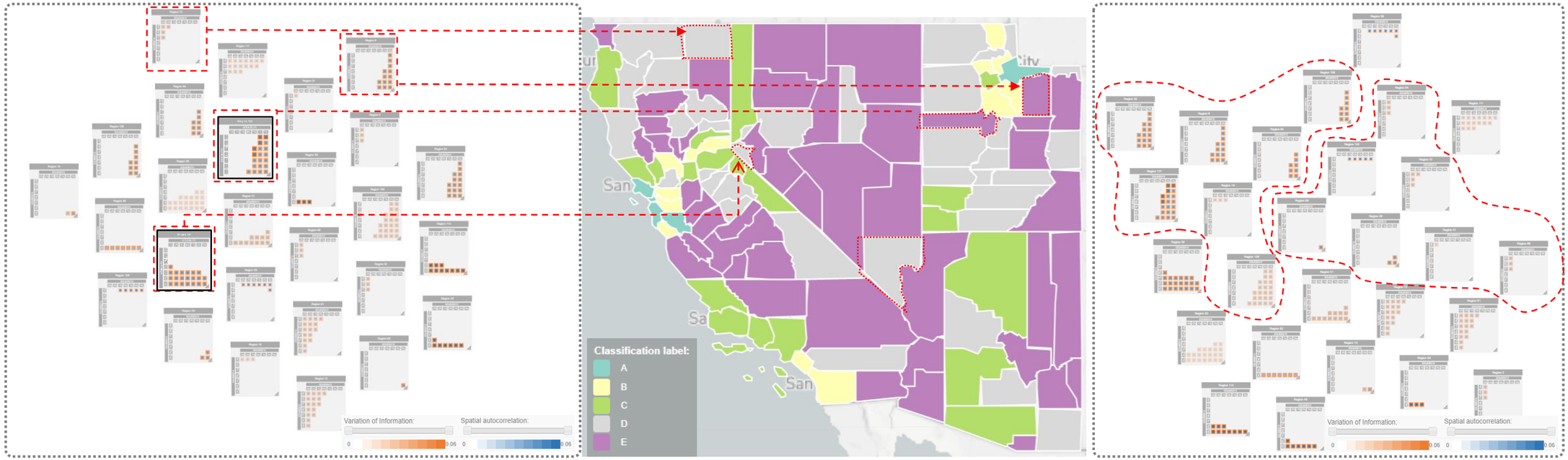




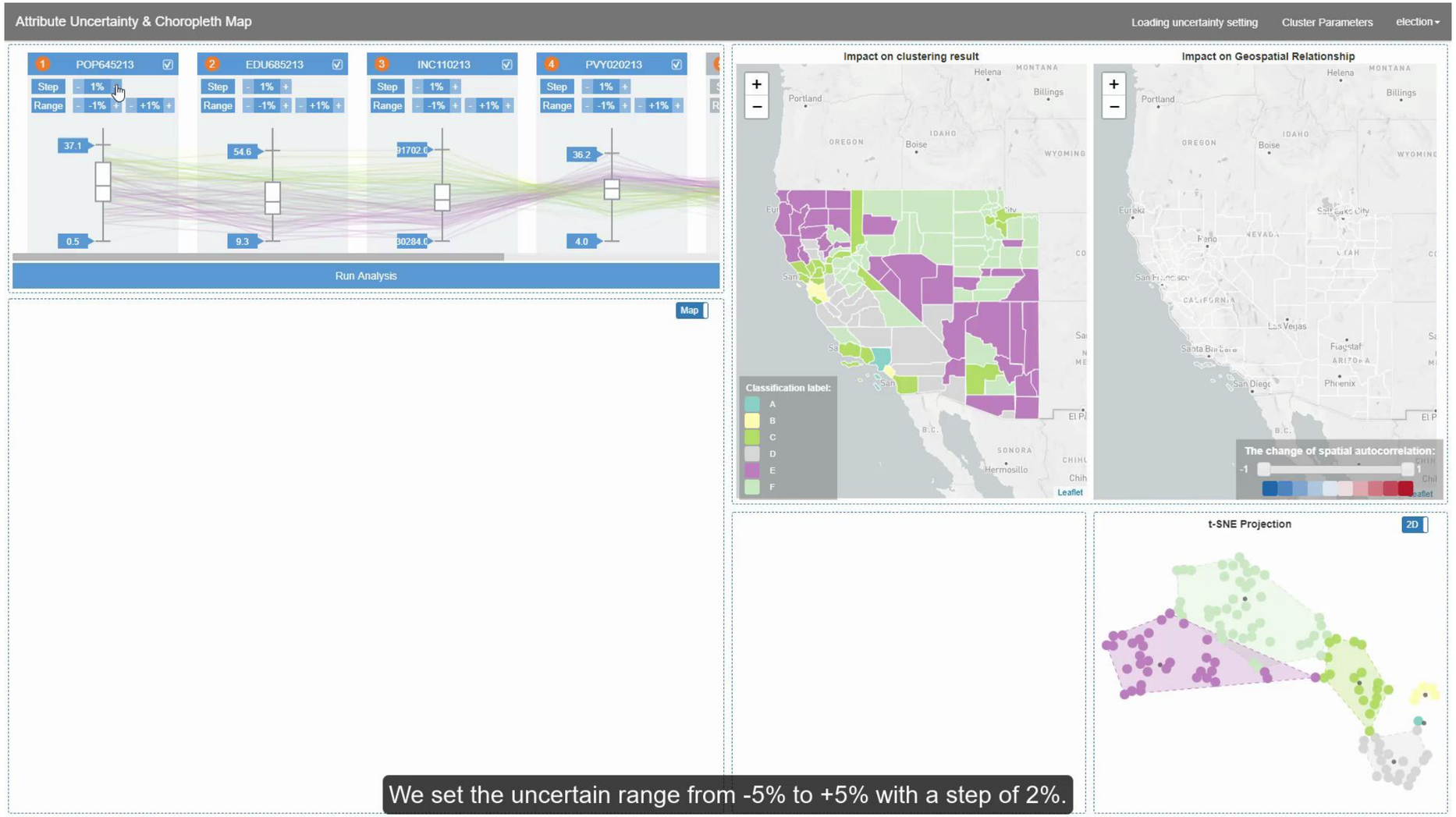
Exploring the Sensitivity of Choropleths under Attribute Uncertainty

Impact Matrix – Multi-Attribute Impact Profile

Map-Based and PCA-Based Layouts

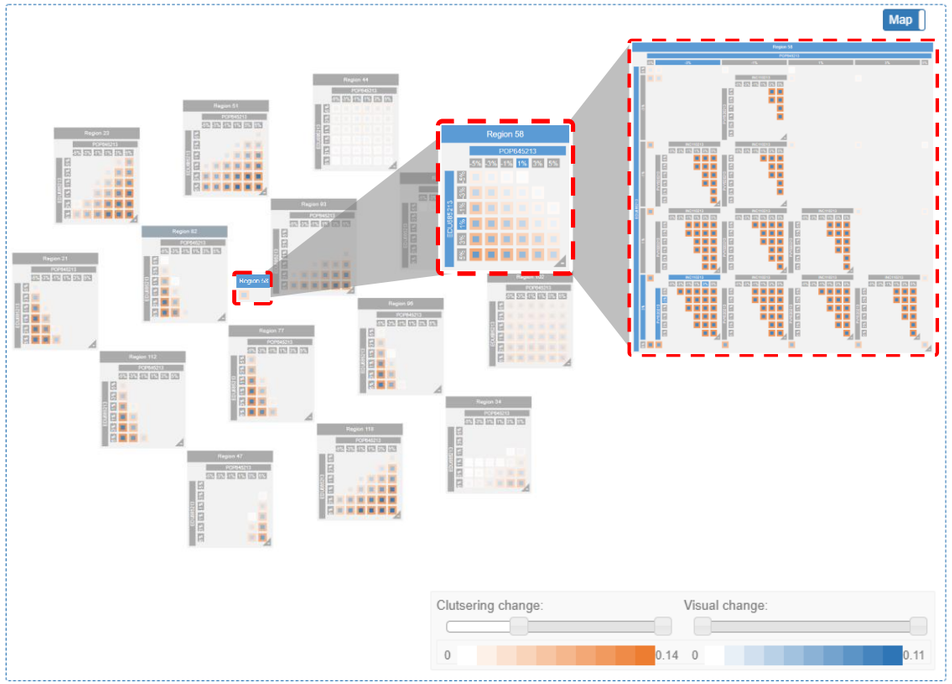


Case study



We set the uncertain range from -5% to +5% with a step of 2%.

Future Work



- To deal with the increased data dimension and the number of spatial units
  - increased computational cost
  - inconvenient to show the entire matrix by clicking the rectangles one by one.
- Some findings and discoveries might also be related to the clustering algorithm.
- To consider cascading effects when multiple units change simultaneously.

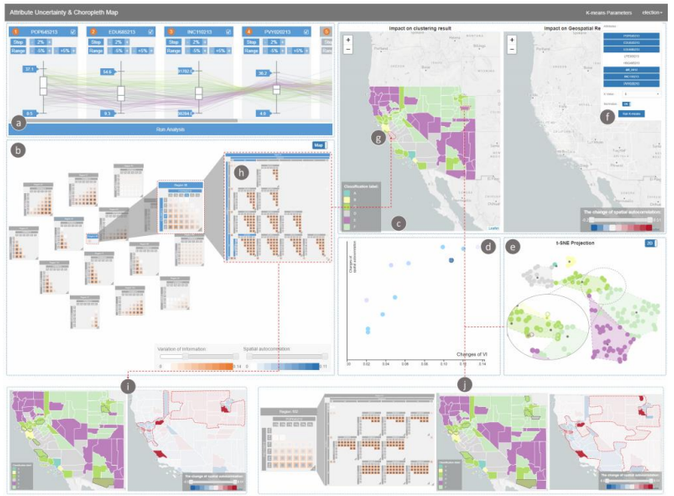
Exploring the Sensitivity of Choropleths under Attribute Uncertainty

Zhaosong Huang , Yafeng Lu , Elizabeth A. Mack , Wei Chen , and Ross Maciejewski

System available at : https://github.com/VADERASU/Choropleths-Attribute-Uncertainty

Thank you!

Q&A



The State Key Lab of CAD&CG, Zhejiang University

The School of Computing, Informatics & DSE, Arizona State University

The Department of Geography, Michigan State University

This work is supported by :

1. The National 973 Program of China (2015CB352503)
2. The National Natural Science Foundation of China (61772456, 61761136020).
3. The National Science Foundation (1350573)
4. The U.S. Department of Homeland Security under Grant Award 2017-ST-061-QA0001.

Z. Huang, Y. Lu, E. Mack, W. Chen, R. Maciejewski, “Exploring the Sensitivity of Choropleths under Attribute Uncertainty,” *IEEE Transactions on Visualization and Computer Graphics*, To Appear.