Spatial Statistics in Practice: When and Why does *Where* matter?

Aishwarya Venkat

Tufts University Research and Data Symposium for Nutrition 2022
March 3, 2022

Food Prices for Nutrition

Tufts University | Gerald J. and Dorothy R. Friedman School of Nutrition Science and Policy

Outline

- Foundations
- Tools of the Trade
- Research example
- Lessons and Resources

Foundations

Spatial Thinking

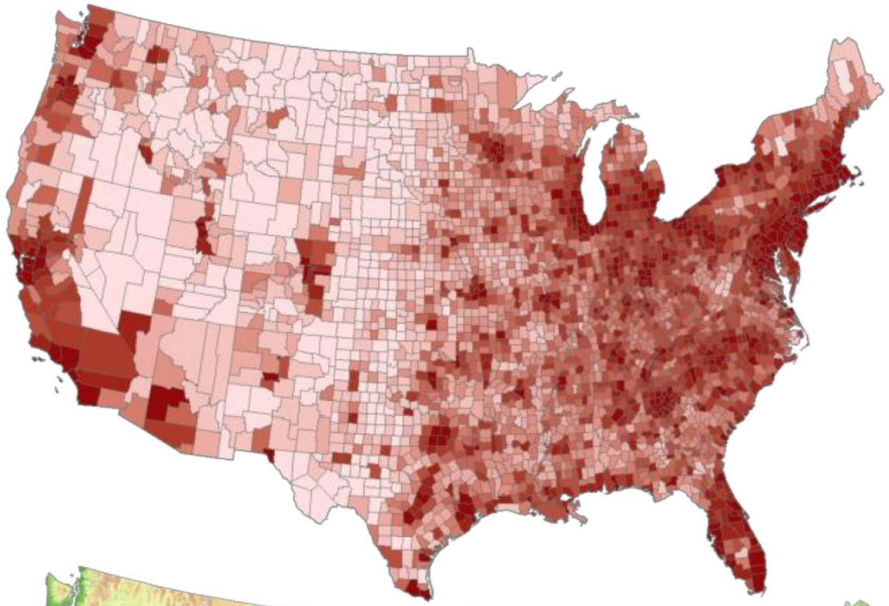
- By studying space, we study:
  - the shape, size, orientation, location, direction or trajectory of objects, processes or phenomena
  - the relative spatial positions of multiple objects, process or phenomena
- Location is a first-level descriptor
  - **Distributions** with patterns, density, and other sample properties
  - **Regions** with similar or different properties
  - **Hierarchies** with multiple nested levels
  - **Networks** of linked features, spatial or otherwise
  - **Surfaces** (3D)

“Everything is related to everything else. But near things are more related than distant things”

- Waldo R. Tobler, 1970

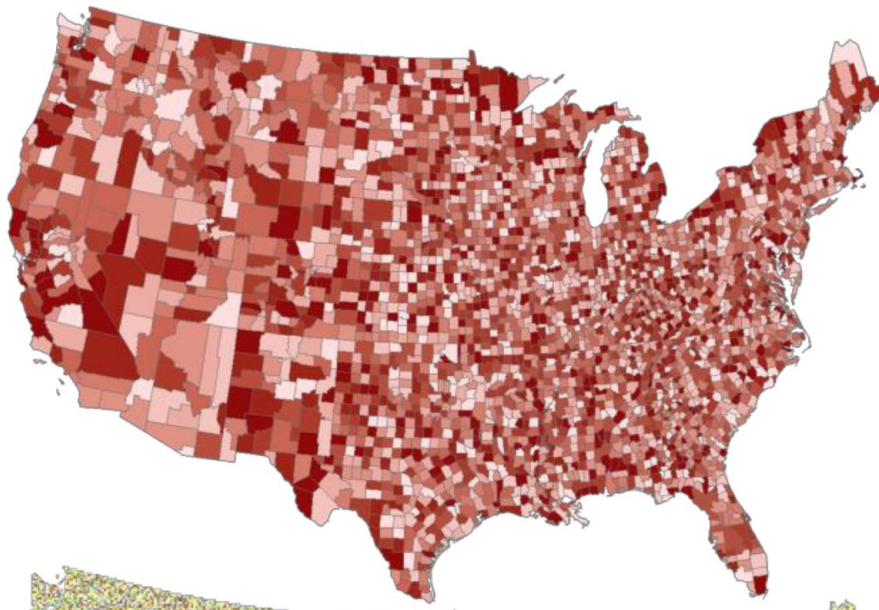
Geographic variables exhibit uncontrolled variance

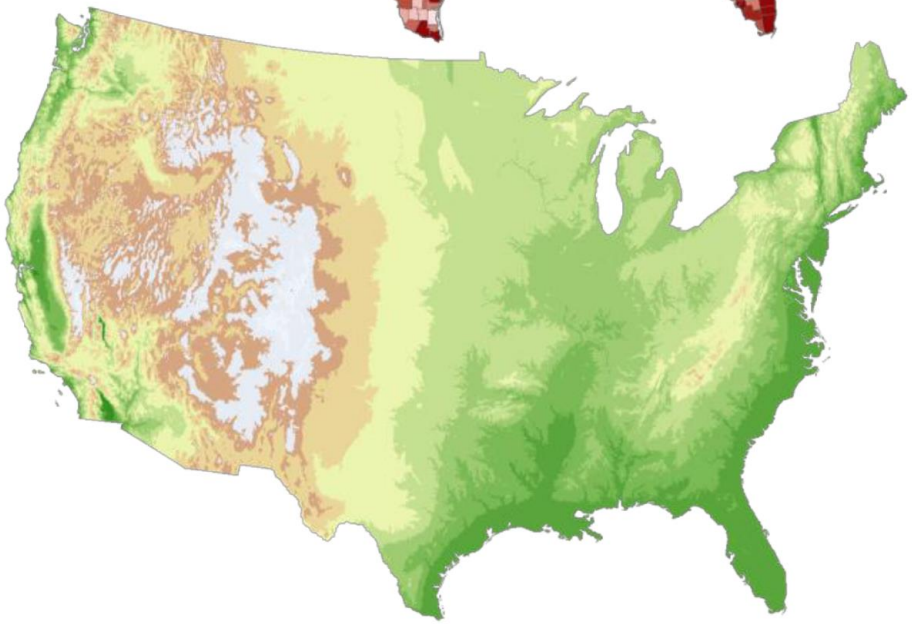
- Michael F. Goodchild, 2004



If features were randomly distributed ...

... population density map of the US would look like this



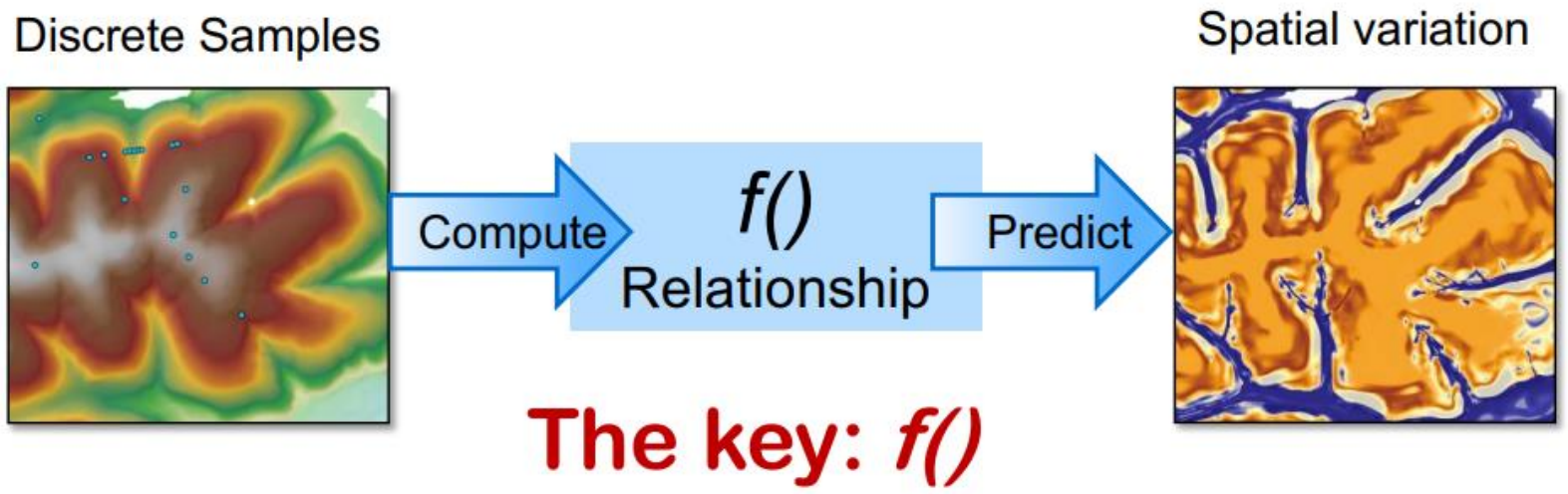


... elevation map of the US would look like this

Source: Gimond, M. Chapter 13: Spatial Autocorrelation. In: Intro to GIS and Spatial Analysis. (2022).

Kinds of Analyses

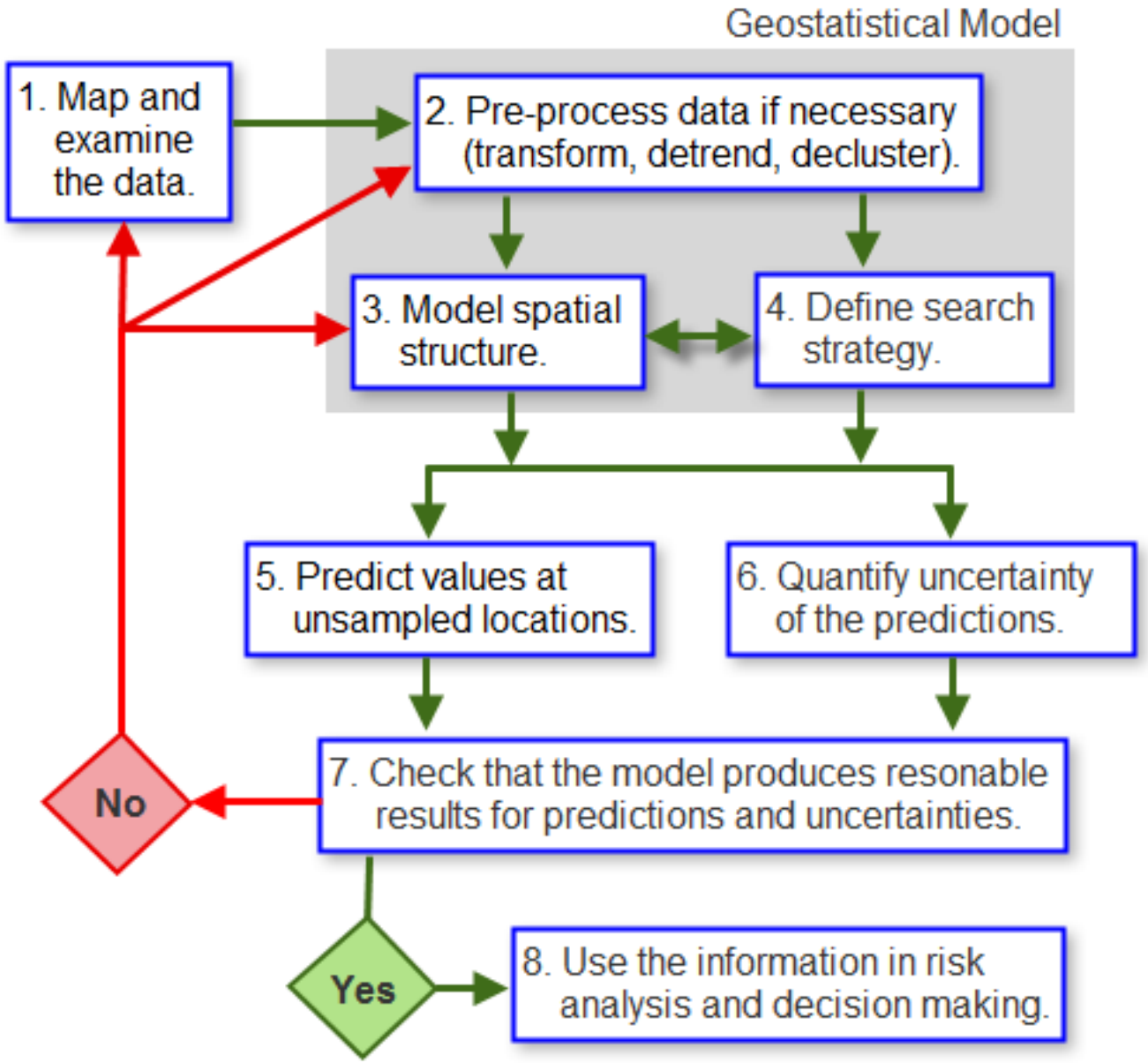
- Descriptive : What happened?
- Diagnostic : Why did it happen?
- Predictive : What will happen?
- Prescriptive : What should happen?



Source: Zhu, A.X. Spatial Prediction and Laws of Geography. (2018).

Workflow

- 0. Collect and extract the **right** data
  - From reputable data sources
  - For the correct spatial and temporal extent

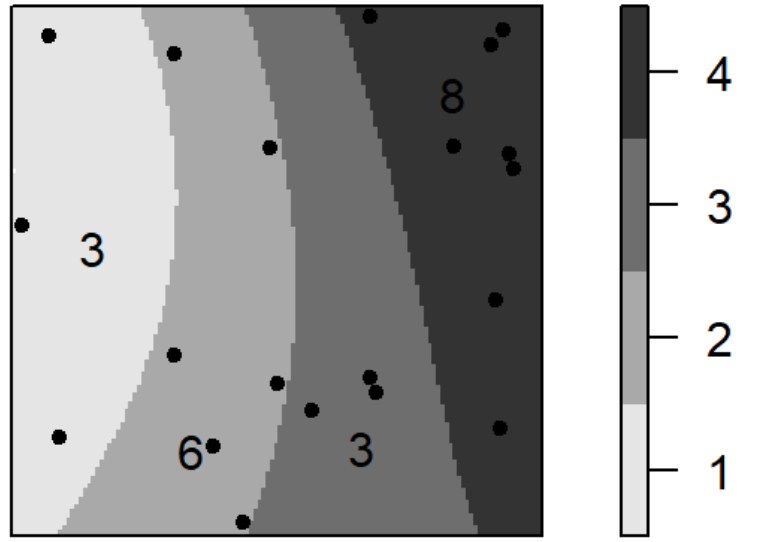


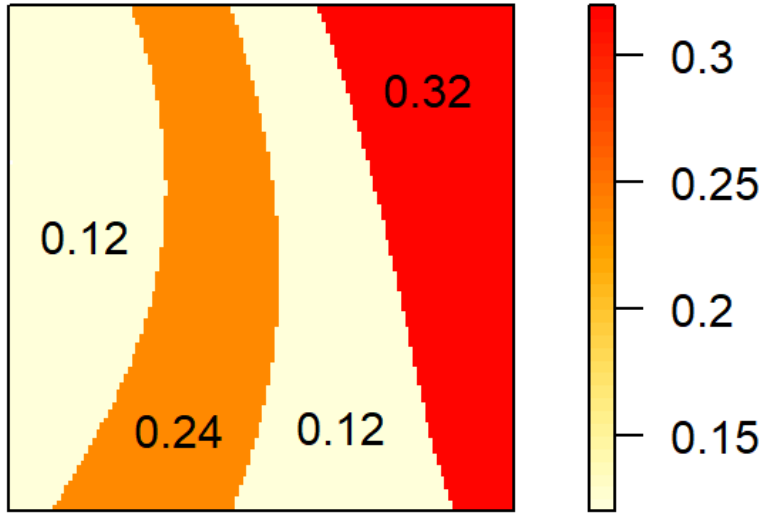
Source: ESRI. ArcGIS Pro. The Geostatistical Workflow.

Key Techniques

Point Pattern Analysis

- Point Pattern: set of events/points in a study region
  - Pattern is realized by an underlying process
  - Useful for points with only categorical descriptors: e.g. trees
- First-order properties: variation over a surface
  - Density, intensity (points/area)
- Second-order properties: observations' influence on one another
  - Average nearest neighbor
  - K, L, g functions

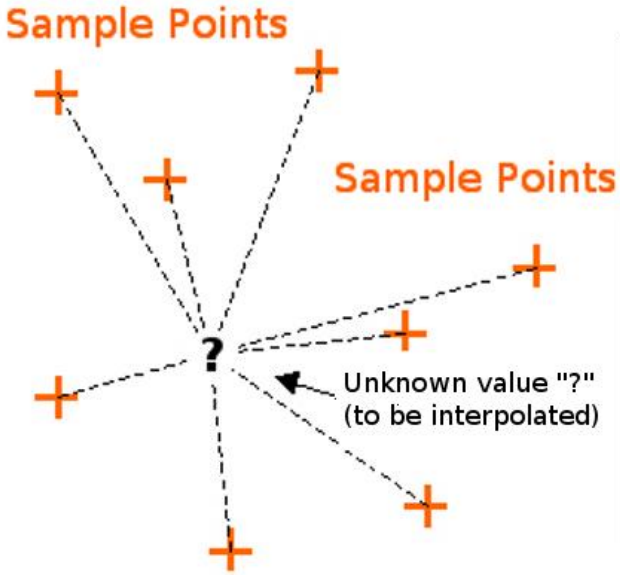


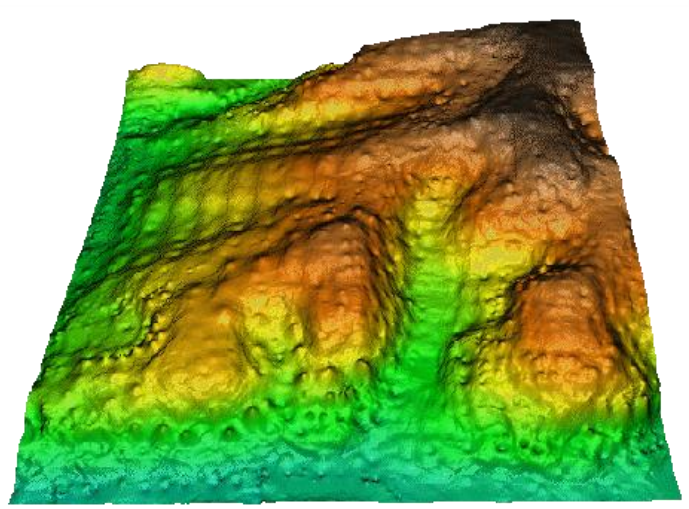


Source: Gimond, M. Chapter 11: Point Pattern Analysis. In: Intro to GIS and Spatial Analysis. (2022).

Interpolation

- “Filling in the blanks” based on sampled locations
- Deterministic (no randomness)
  - Thiessen polygons
  - Density estimation
  - Inverse distance weighted
  - Splines
- Stochastic (some randomness)
  - Kriging

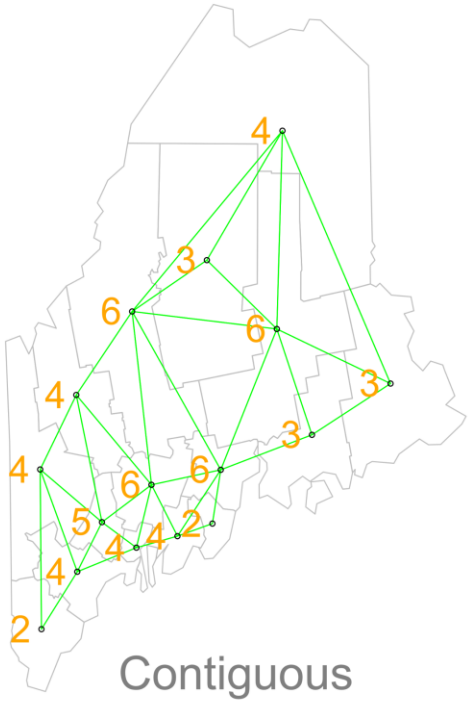


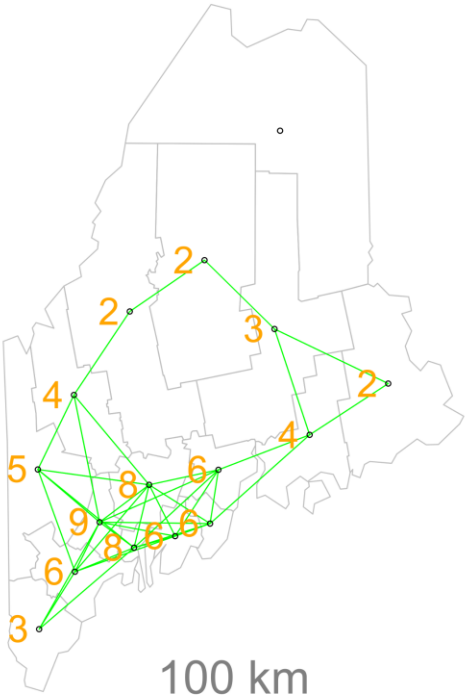


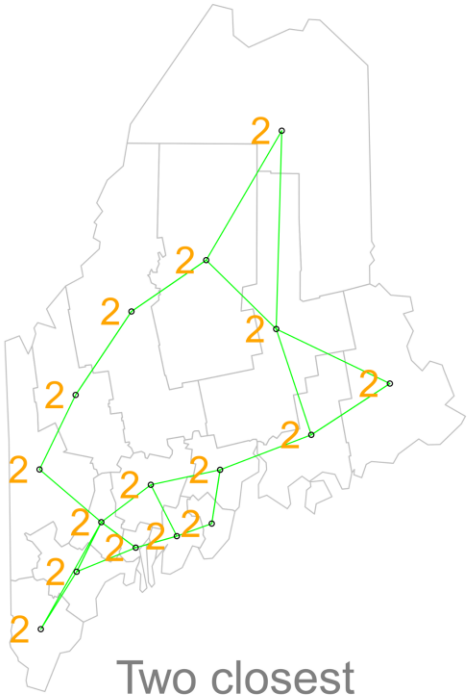
Source: QGIS Team. Spatial Analysis (Interpolation). In: *A Gentle Introduction to GIS*. QGIS v2.18 Documentation (2017).

Spatial Weights Matrix

- Binary (0/1) representation of how features are related to each other







Source: Gimond, M. Chapter 13: Spatial Autocorrelation. In: Intro to GIS and Spatial Analysis. (2022).
Canche, M.S.G. Matrices of Influence Lecture Notes. (2020).

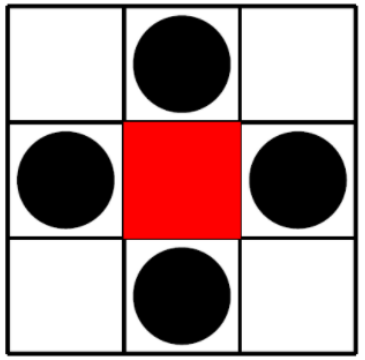


Figure 1: Rook's

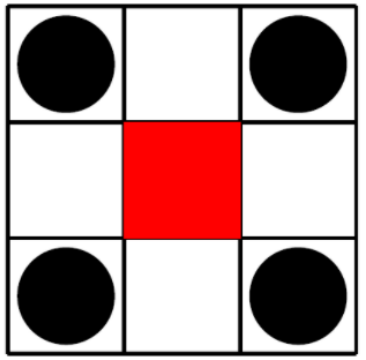


Figure 2: Bishop's

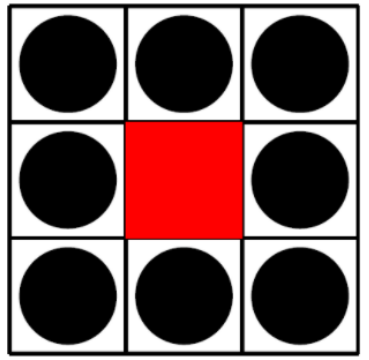
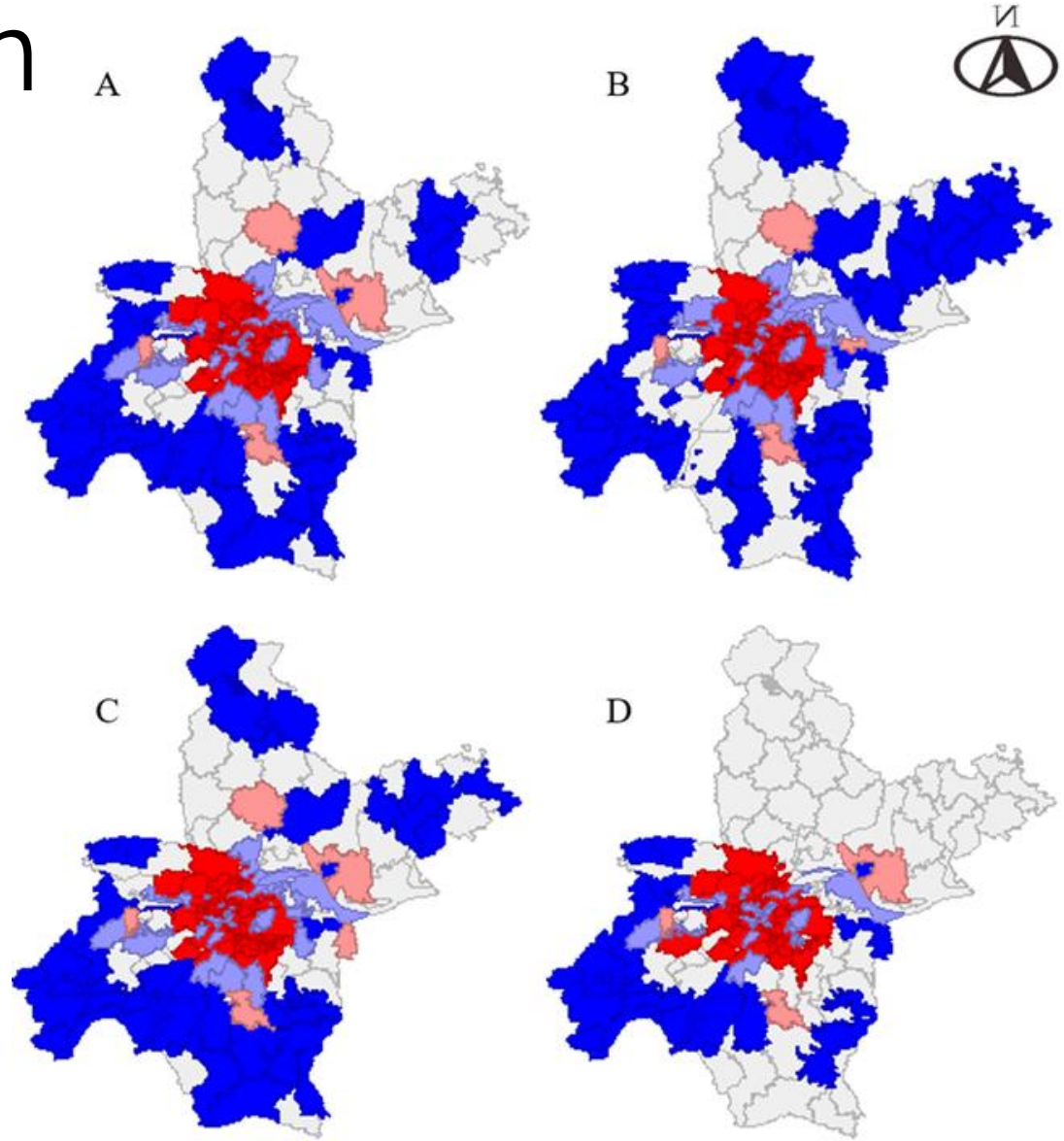


Figure 3: Queen's

Spatial Autocorrelation

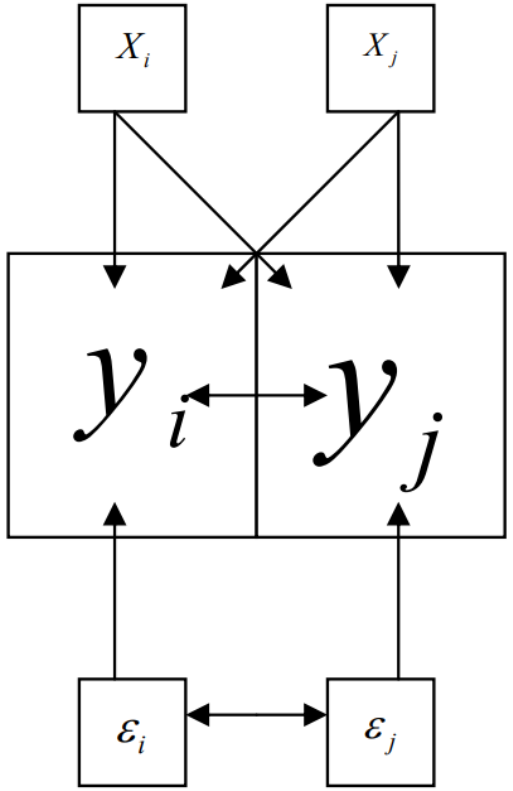
- Degree of correlation between neighboring values
- Metrics: ***Moran's I, Geary's C, Getis-Ord G***
  - **Global:** overall extent of clustering
  - **Local:** clusters and outliers of high and low values
- Local Indicators of Spatial Association (LISA)

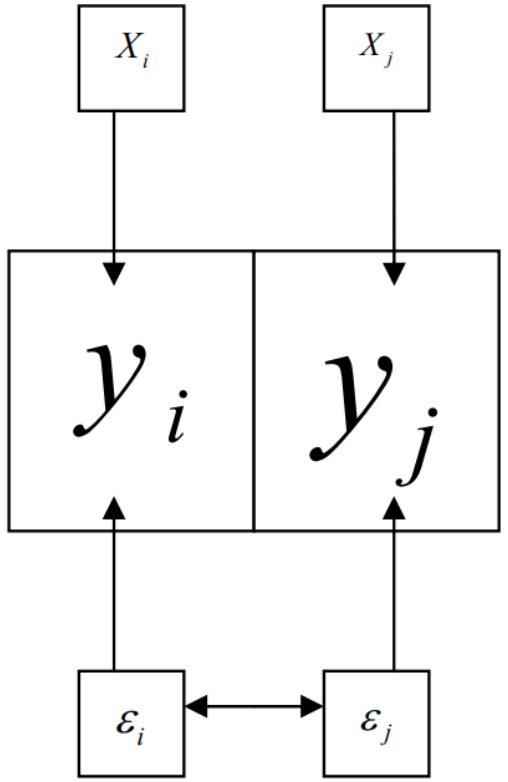


Source: Liu, W., Wang, D., Hua, S., Xie, C., Wang, B., Qiu, W., & Chen, W. (2021). Spatiotemporal analysis of COVID-19 outbreaks in Wuhan, China. Scientific reports, 11(1), 1-9.

Spatial Regressions

- Spatial data often violates normality assumptions for OLS
- **Spatial Lag**: outcome and predictors are correlated in space
- **Spatial Error**: only error terms across different spatial units are correlated
- **Geographically Weighted Regression**: one regression equation for each feature, accounting for its closest neighbors



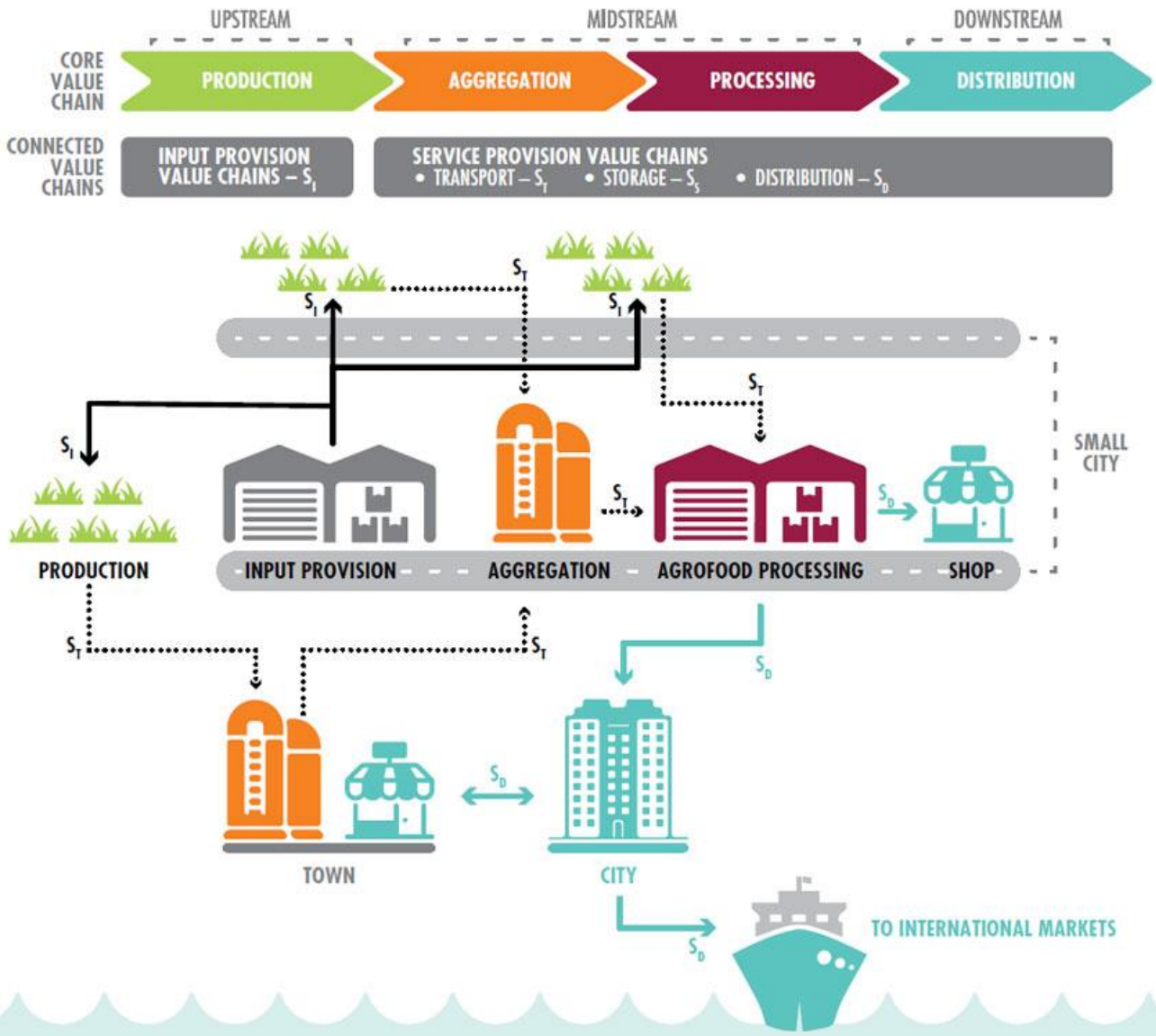


Source: Spatial Regressions with Geoda Training Module. Spatial Structures in the Social Sciences, Brown University.

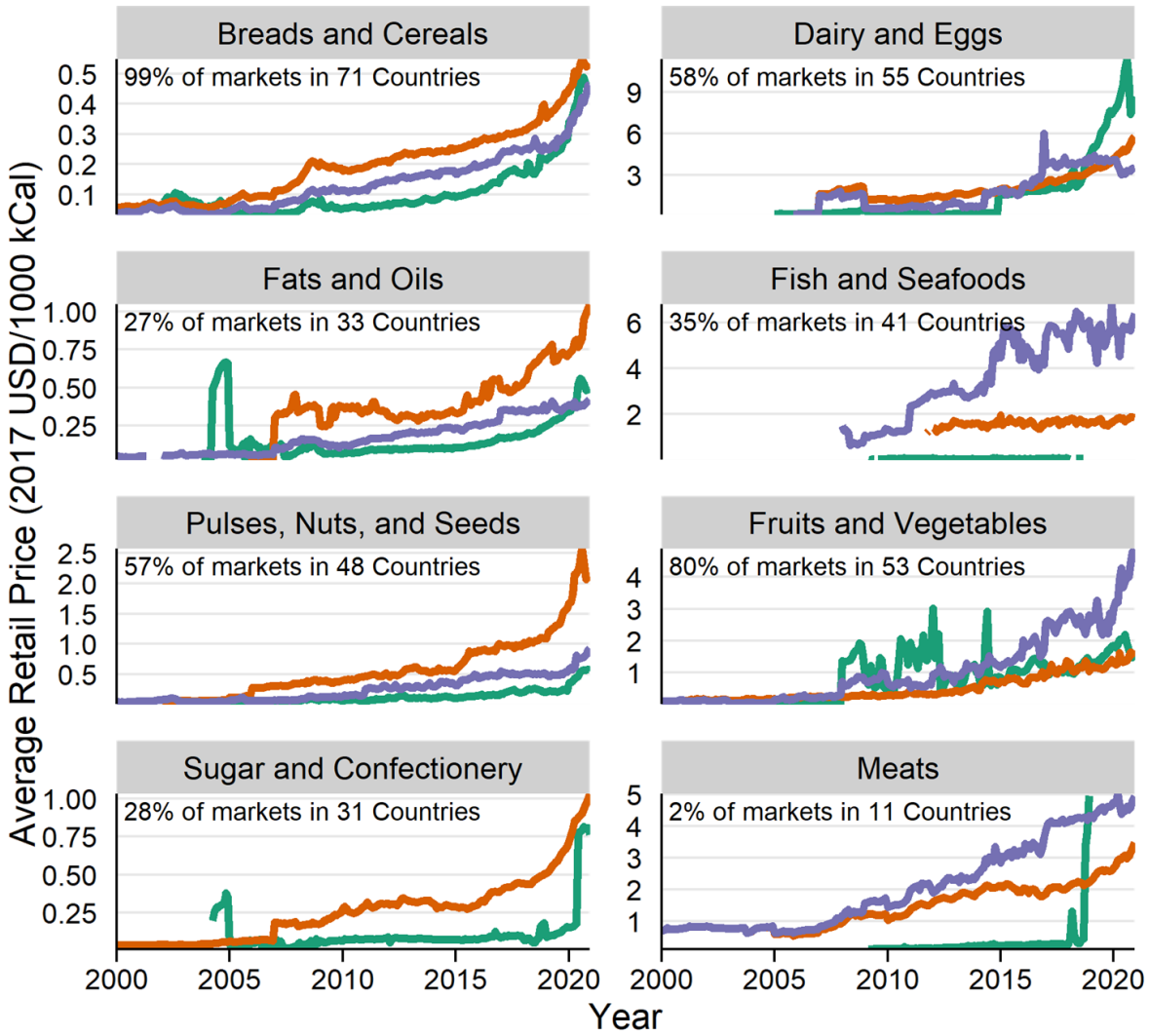
Research Example

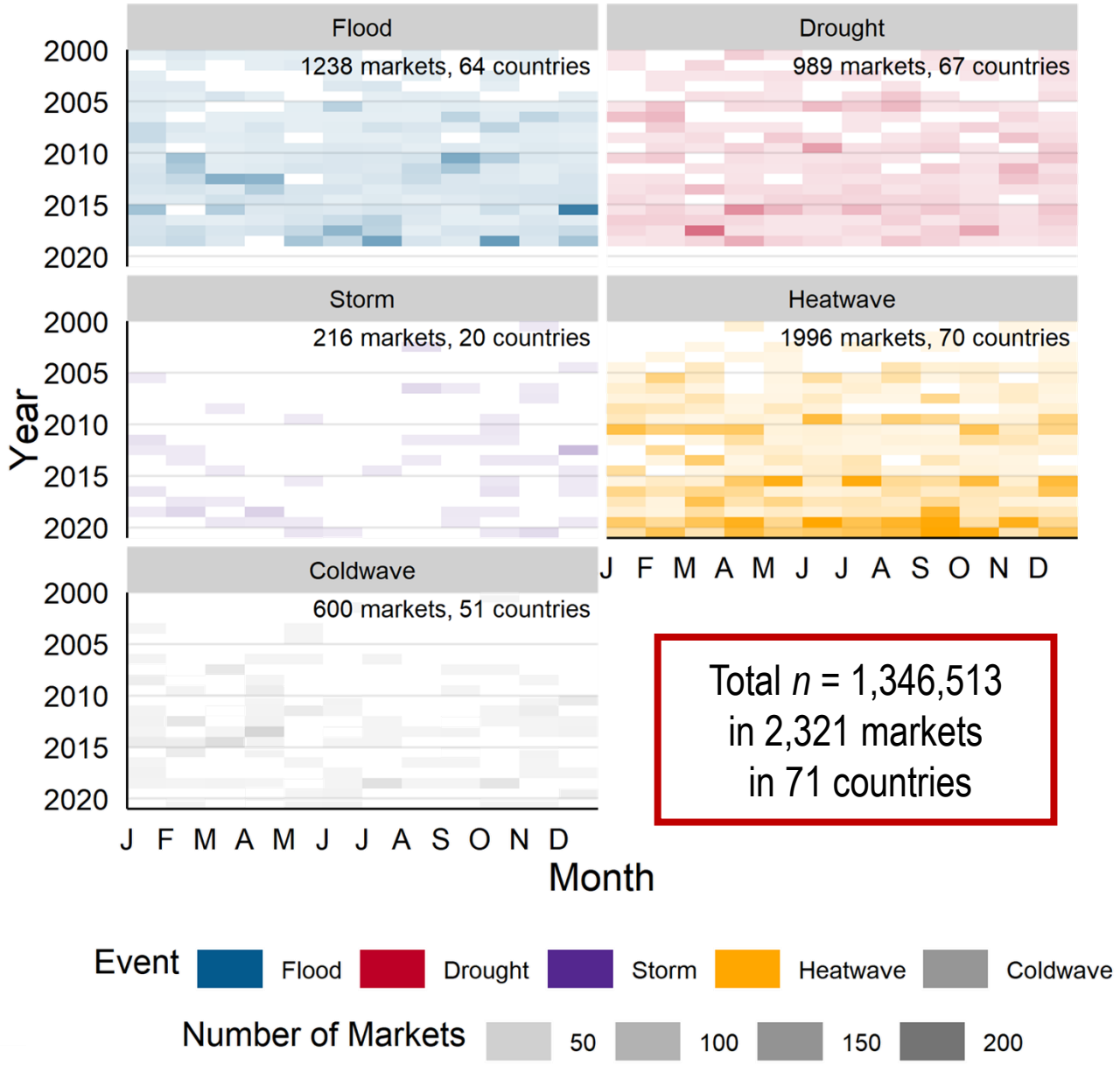
Research Questions

- Big question: do extreme weather events affect retail food prices?
- Spatial questions
  - How integrated are markets?
  - To what degree is price in one market affected by price in:
    - nearest city
    - nearest port city
    - rest of world
  - Does observed effect persist after accounting for spatial lag?

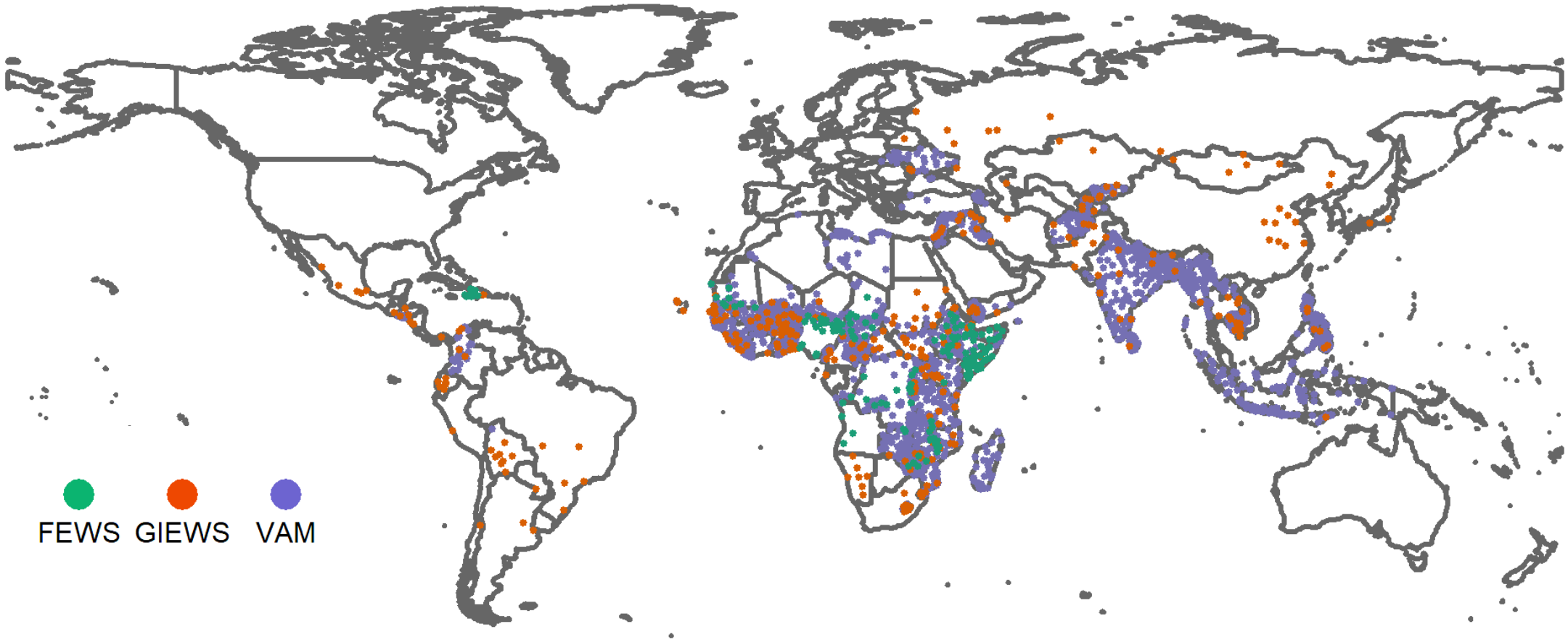


Source: Figure 12 in *The State of Food and Agriculture 2017*: Leveraging Food Systems for Inclusive Rural Transformation. 2017. Rome, FAO.


Breads and Cereals
99% of markets in 71 Countries
Dairy and Eggs
58% of markets in 55 Countries
Fats and Oils
27% of markets in 33 Countries
Fish and Seafoods
35% of markets in 41 Countries
Pulses, Nuts, and Seeds
57% of markets in 48 Countries
Fruits and Vegetables
80% of markets in 53 Countries
Sugar and Confectionery
28% of markets in 31 Countries
Meats
2% of markets in 11 Countries
Average Retail Price (2017 USD/1000 kCal)
Year
Dataset
FEWS
GIEWS
VAM

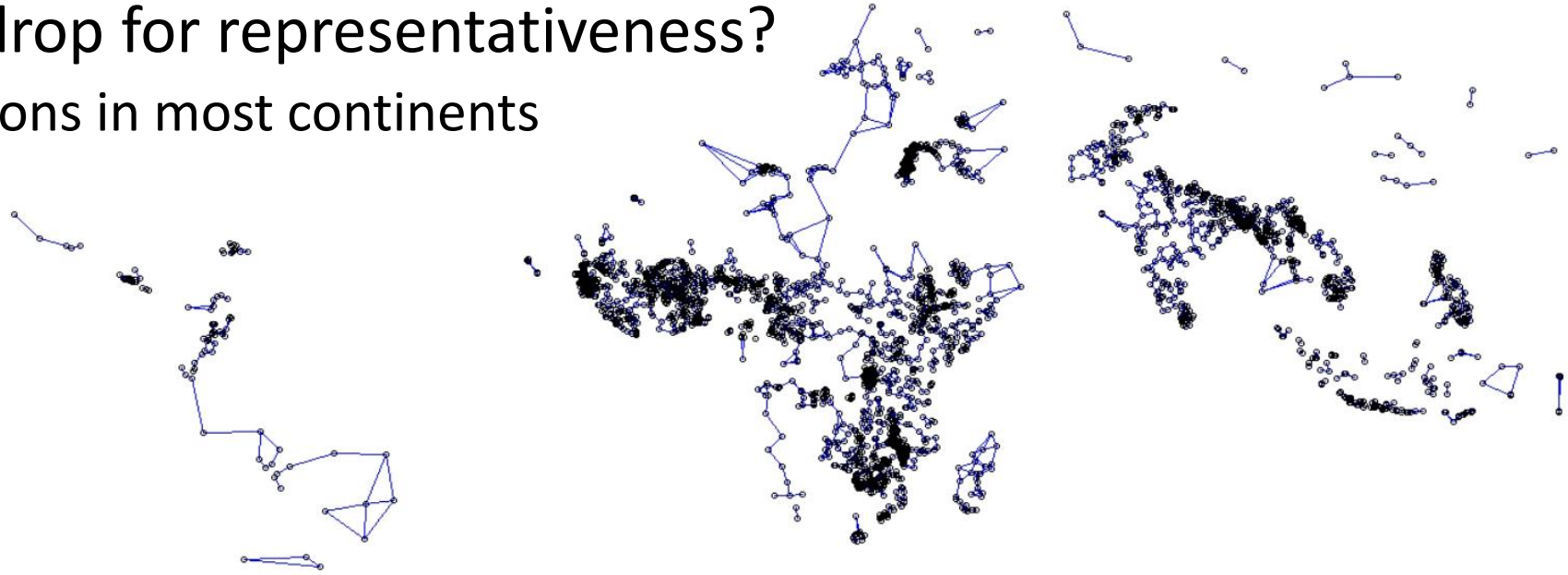

Flood
1238 markets, 64 countries
Drought
989 markets, 67 countries
Storm
216 markets, 20 countries
Heatwave
1996 markets, 70 countries
Coldwave
600 markets, 51 countries
Year
Month
Total n = 1,346,513
in 2,321 markets
in 71 countries
Event
Flood
Drought
Storm
Heatwave
Coldwave
Number of Markets
50
100
150
200

Markets



Spatial Decisions

- Is there an underlying process?
  - Yes! Retail price is observed at some markets in this dataset but similar price dynamics hold true in unobserved markets
- How many neighbors to consider, at what distance?
- Which markets to drop for representativeness?
  - Very few observations in most continents



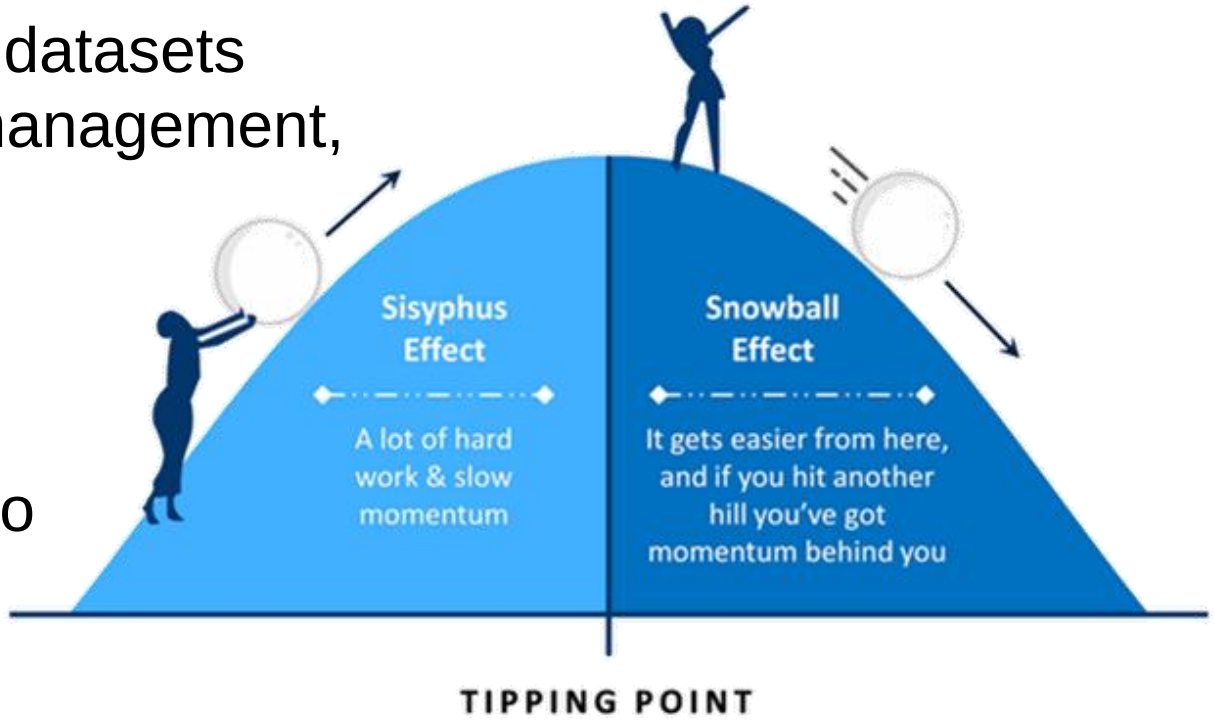
Next steps

- Wholesale – Retail market pairs to study markups
  - Distance-based thresholds—how far can a food item realistically be transported?
  - Travel time
- Changing combination of market/item/price for each time point
- Temporal lags between price at different markets

Lessons and Resources

Lessons

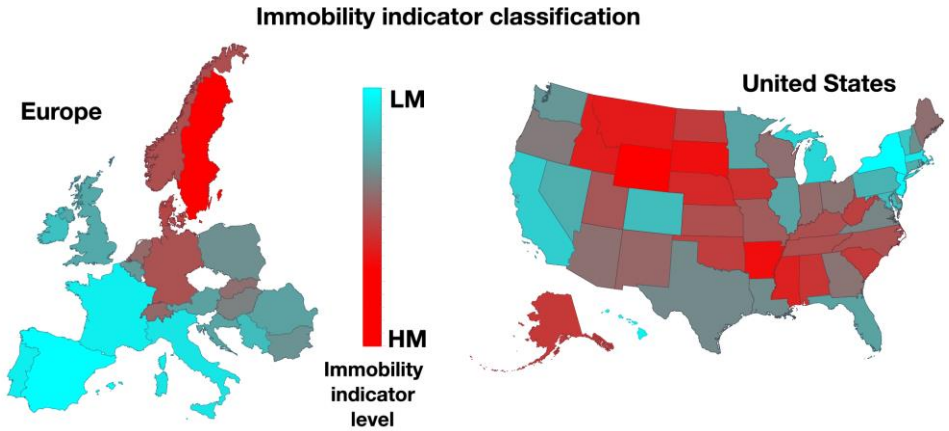
- Identify the universe of predictors early
  - Everything is related, but not everything is relevant/directly measurable
- Research techniques
  - Explore and bookmark interesting datasets
  - Related skills: data mining, data management, independent learning
  - Learn by doing, snowball effect!
- Logistics
  - Workflow matters!
  - Document data decisions on the go
  - Keep reproducibility in mind

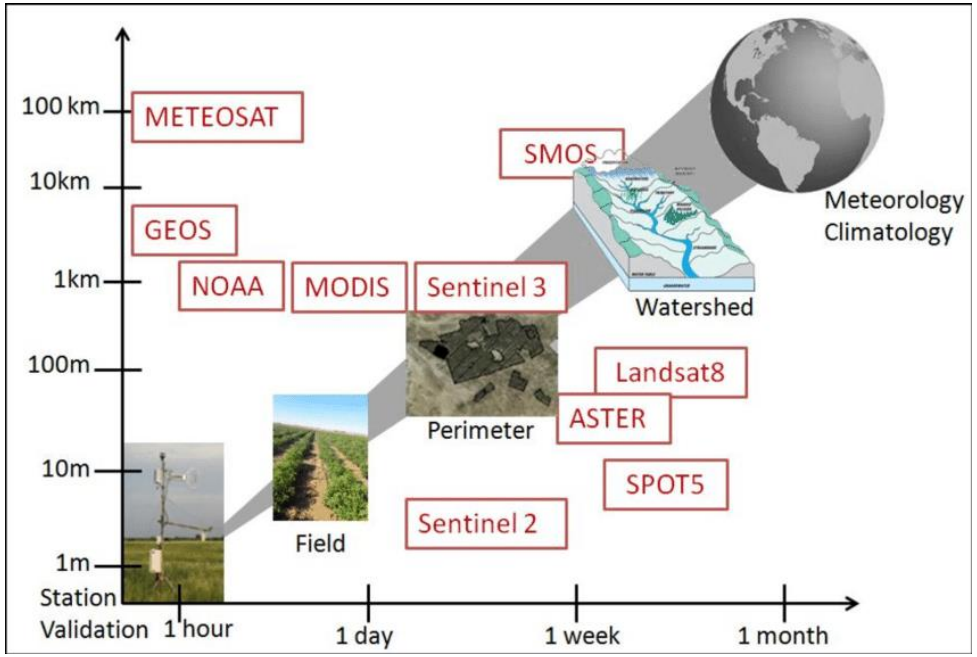


Source: SketchBubble

New Data & New Demands

- Spatial data is becoming ubiquitous
  - At multiple spatial scales (local, regional, national, global)
  - At high resolution
- User-generated data
  - Auxiliary, unstructured, individual





Sources: Malbéteau, Y. (2016). Suivi des ressources en eau par une approche combinant la télédétection multi-capteur et la modélisation phénoménologique (Doctoral dissertation, Université Paul Sabatier-Toulouse III).

Cot, C., Cacciapaglia, G., & Sannino, F. (2021). Mining Google and Apple mobility data: temporal anatomy for COVID-19 social distancing. Scientific reports, 11(1), 1-8.

Resources

- R
  - RSpatial
  - Spatial Regression Analysis in R: A Workbook
  - Geocomputation with R
- Python
  - Geographic Data Science with Python
  - PySAL
  - Spatial Modelling for Data Scientists
- Stata
  - Reference Manual: Spatial Autoregressive Models
  - Lecture Notes: Raschky, Drukker, Pisati

Thank you!
Questions?

Aishwarya Venkat
Aishwarya.Venkat@tufts.edu