Bootstrap Methods for Inference with Spatial Panel Data Models

Drake Warren

University of Illinois at Urbana-Champaign

Regional Economics and Public Policy Research Group (REAP)

Outline

- Motivation behind spatial panel bootstrap methods
- Spatial Window Block Bootstrap Method for Spatial Panel Data (SWBB)
- Spatial Residual Block Bootstrap Method for Spatial Panel Data (SRBB)
- Problems with spatial block bootstrap methods when using real data

Motivation

- Very common to use economic impact studies that forecast economic effects using input-output models
  - Essentially cheerleading
  - Always see a positive impact
  - Many tenuous assumptions
- Prefer post hoc analysis
  - Looks at realized effects
  - Fewer assumptions

Motivation

- Need better methods for spatial panel data
  - Methods like multi-period differences in differences often ignore serial correlation and spatial autocorrelation
- Use methods and models that minimize assumptions and let the data speak
  - Quasi-experimental control groups
  - Bootstrap methods

Bootstrapping

- Way of conducting statistical inference
- Assumes that empirical distribution of observed data is representative of true distribution
  - Draw observations from empirical distribution with replacement

Bootstrapping

- Avoids parametric computations
  - Repeatedly calculate statistics of sampled data
  - Analyze distribution of statistics
    - Distributions approximate true distributions
  - Does not make assumptions such as normality
- Examples:
  - Median
  - Inference

Bootstrapping

- More complex sampling methods account for dependence of data
  - Block bootstrapping for panel data
    - Draws all/many periods together
    - Maintains the serial correlation structure
  - Residual resampling for regression models
    - Randomly resamples residuals and re-estimates
  - Spatial block bootstrapping for spatial cross-sectional data
    - Samples spatial blocks

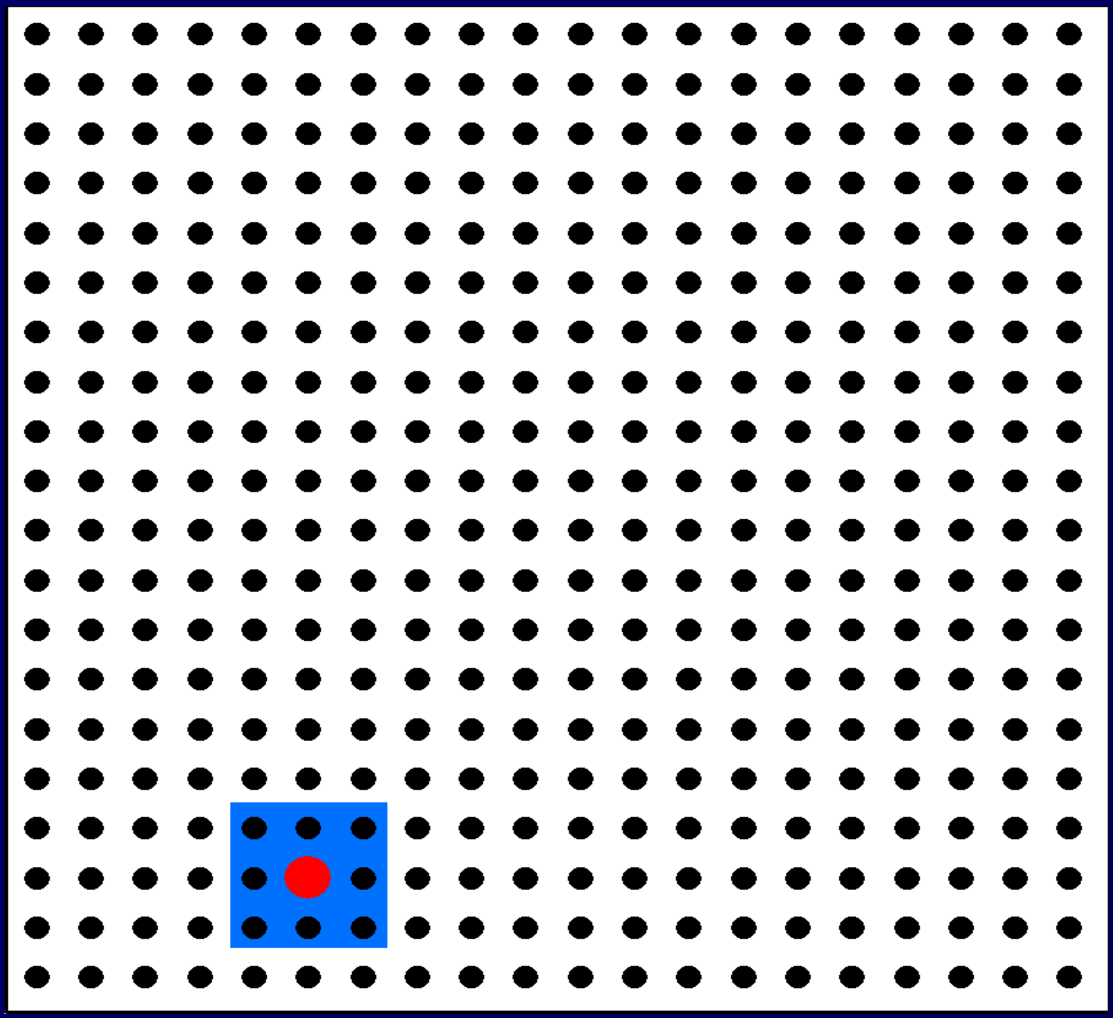
Spatial Window Block Bootstrap

- Modifies spatial window bootstrap for time
- Samples “windows” of all time periods of observations
- Not favored due theoretical considerations (edge effects)
  - Evaluate the severity of problem using Monte Carlo experiments

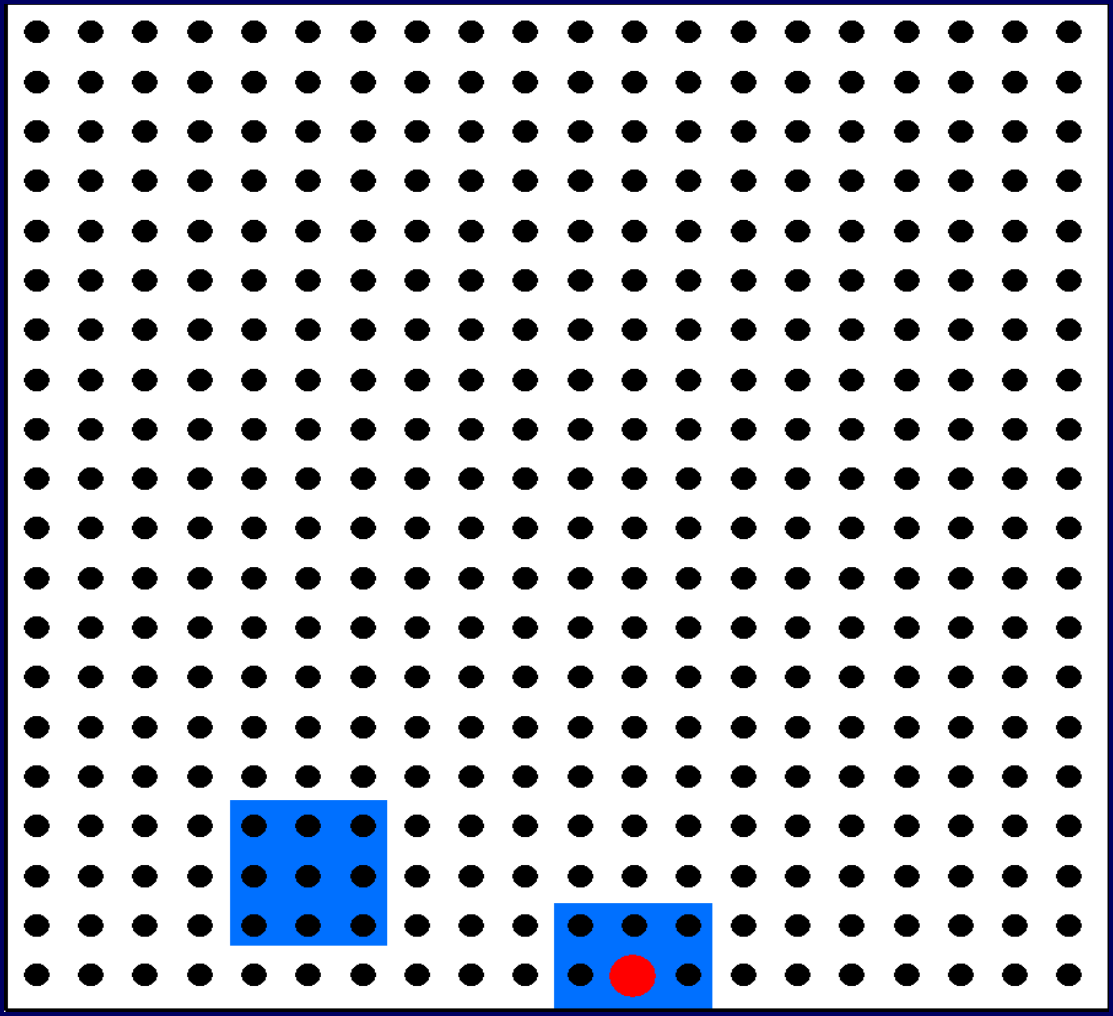
Spatial Window Block Bootstrap

1. Randomly select an observation giving equal chance to all observations. Include all time periods of that observation in pseudo-dataset.
2. Using spatial neighbors information, e.g., spatial weights matrix, include all years of all neighbors.
3. Repeat, with replacement, until number of observations reaches the sample size.

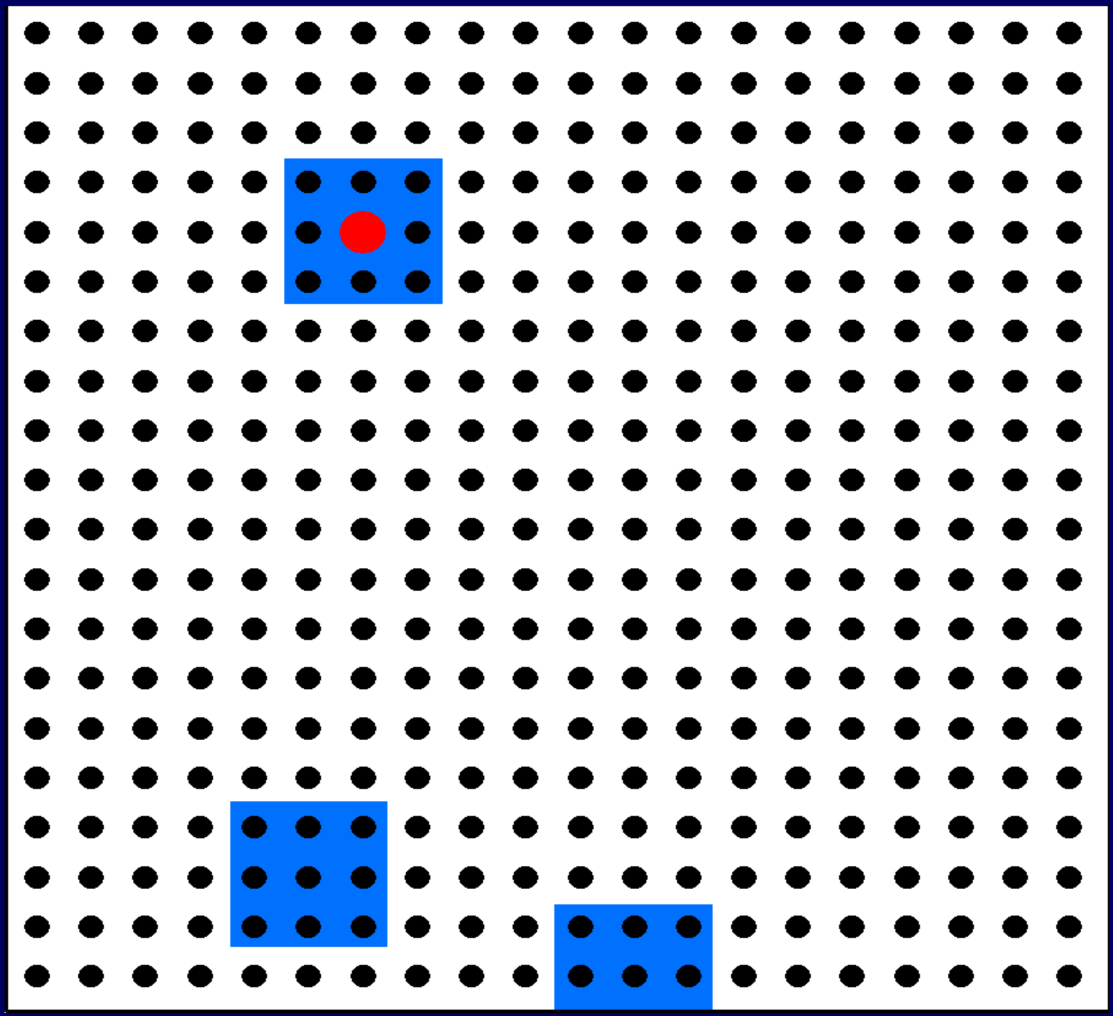
Spatial Window Block Bootstrap



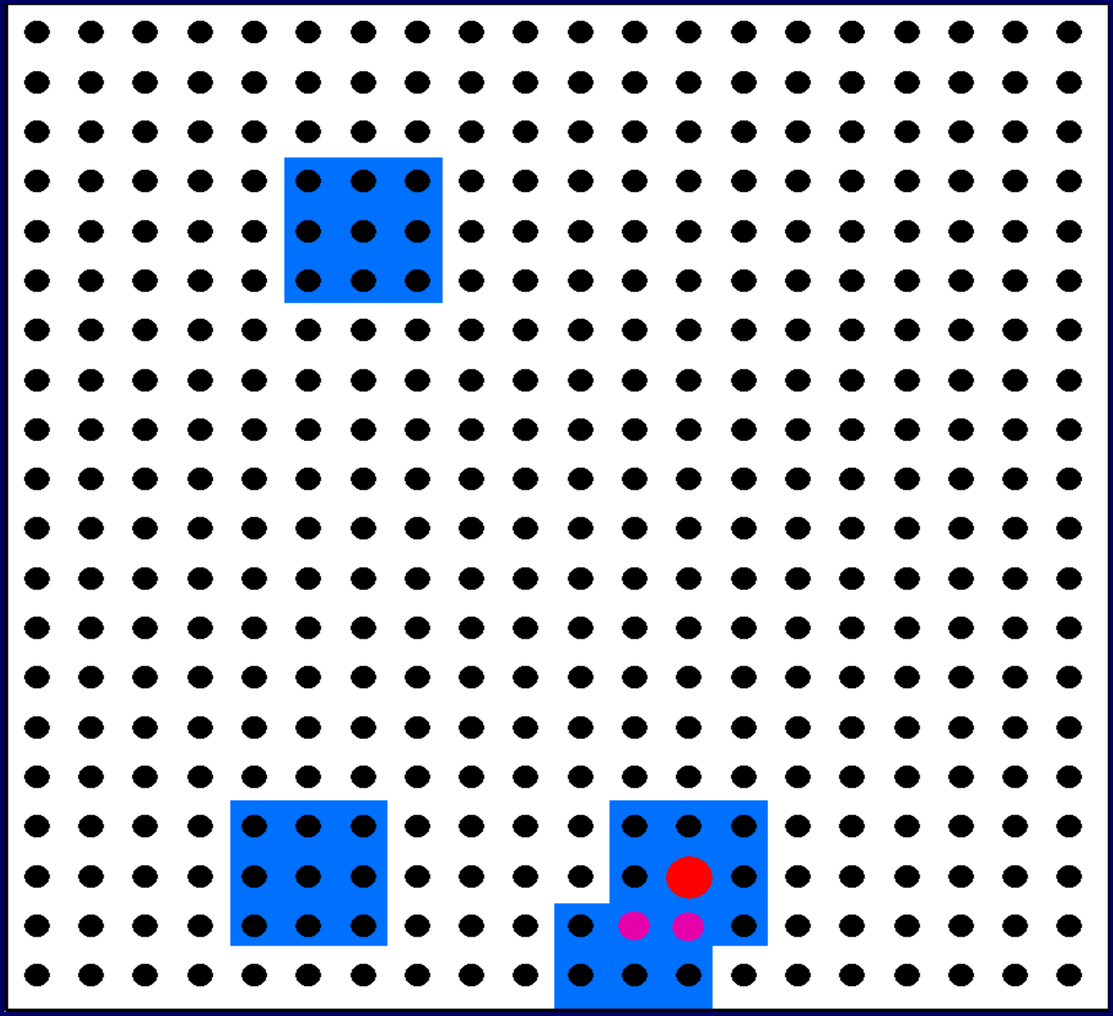
Spatial Window Block Bootstrap



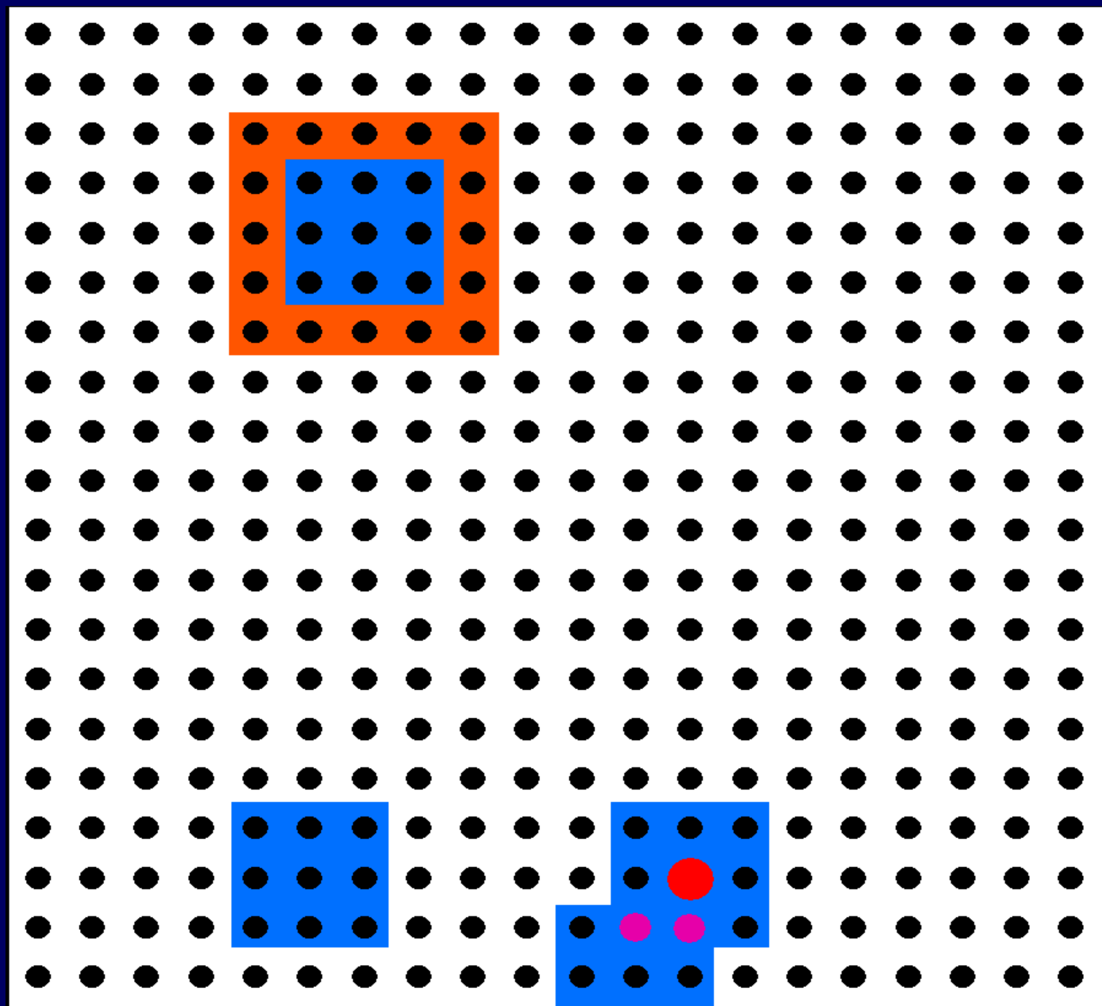
Spatial Window Block Bootstrap



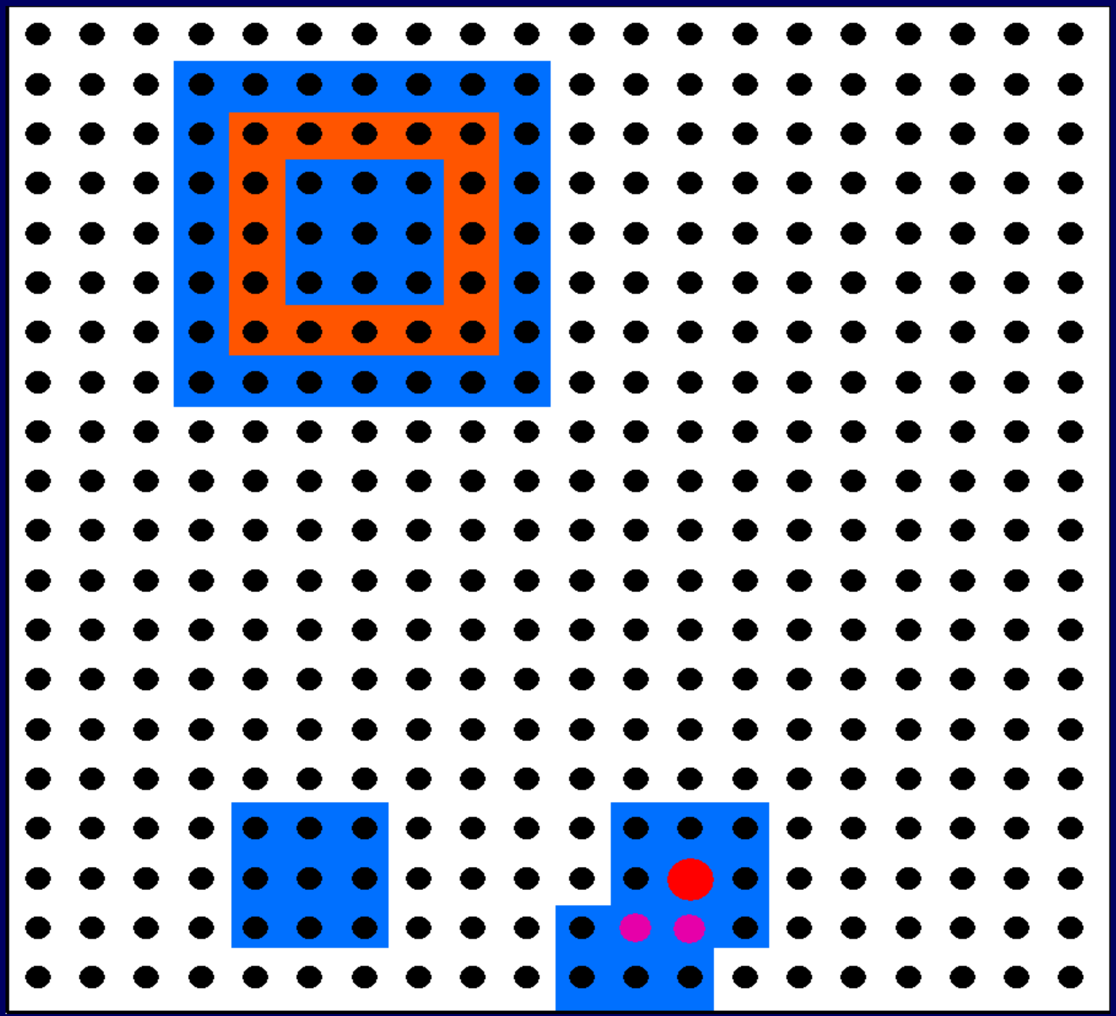
Spatial Window Block Bootstrap



Spatial Window Block Bootstrap



Spatial Window Block Bootstrap



Spatial Window Block Bootstrap

4. Estimate spatial regression using 2SLS
   - Faster than MLE
   - Avoids assumptions of MLE
5. Keep track of estimates
6. Once a large number of simulations are run, analyze distributions of estimates

Spatial Residual Block Bootstrap

1. Estimate the spatial lag coefficient, ρ, using 2SLS with all observed data
   - Spatial lag coefficient needed to adjust residuals
   - Save all residuals
2. Sample, with replacement, all periods of residuals for each selected observation
   - Does not require spatial neighbors information (but spatial lag model does)

Spatial Residual Block Bootstrap

3. Once residuals for all observations have been drawn, adjust dependent variable using the newly assigned residual:

$$y'_t = y_t + (I - \rho W)^{-1}(e'_t - e_t)\forall t$$

4. Estimate the model keeping track of the estimates.
5. Repeat until a large number of bootstraps are run, and analyze distributions.

Monte Carlo Experiments

- Experiments generate data that will be used to compare methods
  - Simulates spatial processes that may have serial correlation, spatial error autocorrelation, and heteroskedasticity
- See how estimates compare to the true parameters used to generate data.

Monte Carlo Experiments

- Compared several different methods:
  - Three orders of windows for SWBB
  - SRBB
  - OLS
  - 2SLS
- Compared several measures:
  - Type I error for β_1
  - Power when $\beta_1 = 0.1$
  - Many others in dissertation

Monte Carlo Experiments

- 500 pseudo dataset generations
- 200 bootstrap simulations for each data generation
  - Lower than it should be, but there are 10,000 total runs
- Approximately 1,300 hours on dual processor

Monte Carlo Experiments: Results

- Heteroskedasticity:
  - SWBB has largest confidence intervals and smallest power, but only acceptable Type I error
    - But, SWBB tends to perform worse with a small grid

Monte Carlo Experiments: Results

- Serial correlation:
  - Type I error is large for OLS and 2SLS
  - SWBB performs well, and SRBB sometimes performs well

Monte Carlo Experiments: Results

- Spatial error autocorrelation:
  - Averages Case:
    - SRBB performs better than SWBB
    - OLS and 2SLS perform well
  - *WX* Instruments Levels Case:
    - OLS and 2SLS perform best with large spatial lag
  - $(I\text{-}\lambda W)^{-1}X$ Instruments Levels Case
    - SWBB performs best with large spatial lag

Monte Carlo Experiments: Results

- Overall conclusions:
  - SWBB performs best in averages cases
    - SRBB performs well, except with heteroskedasticity
  - SWBB performs best in levels cases, provided that good instruments are used
    - SRBB is often useless
  - Methods are weak in the areas where they are expected to be deficient

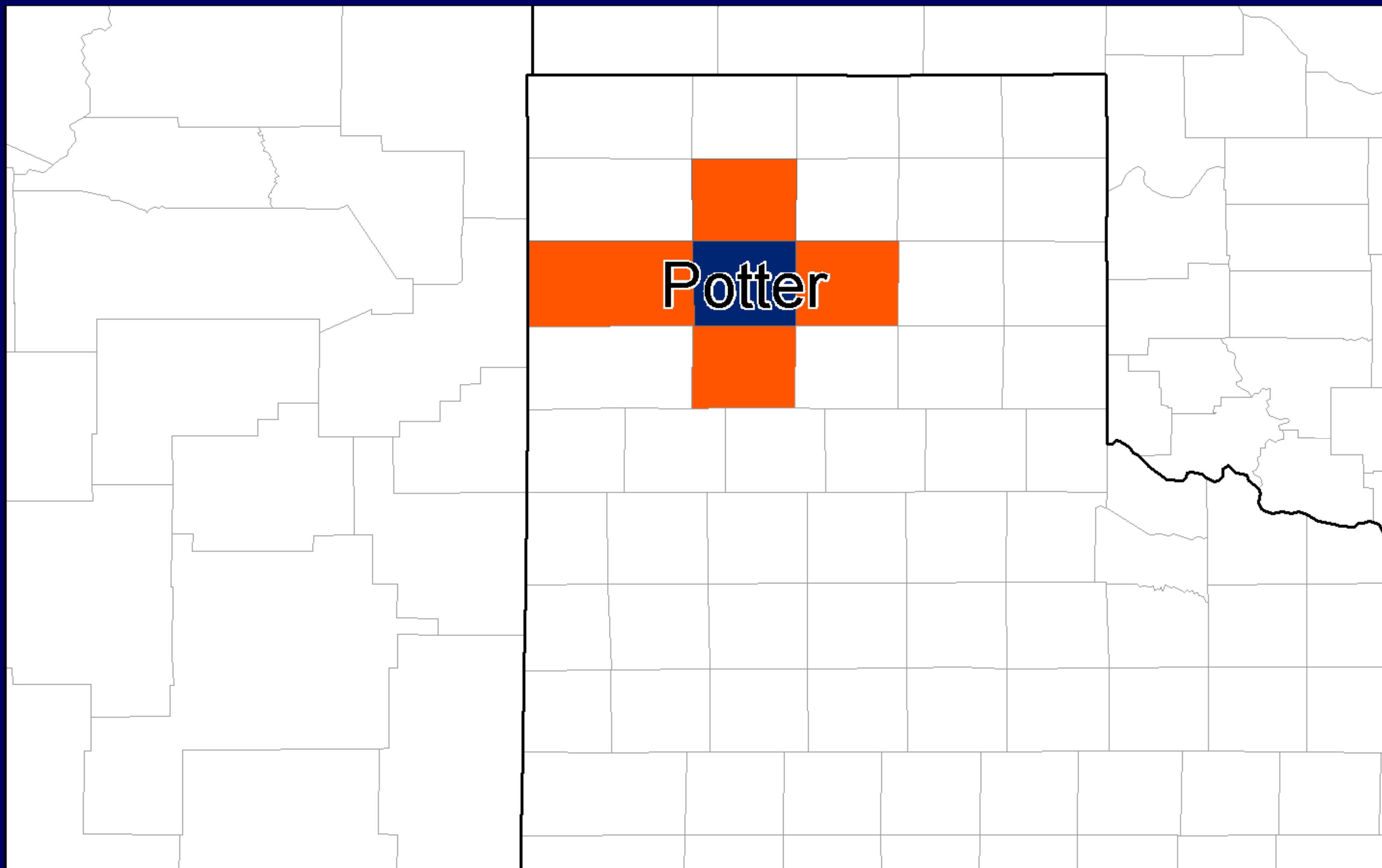
Real Life Problems

- SRBB is limited in its use
  - Cannot handle heteroskedasticity (as seen in Monte Carlo experiments)
  - Cannot be used to estimate random effects models
  - Used only for spatial lag models
    - Performs poorly even with small lag coefficients

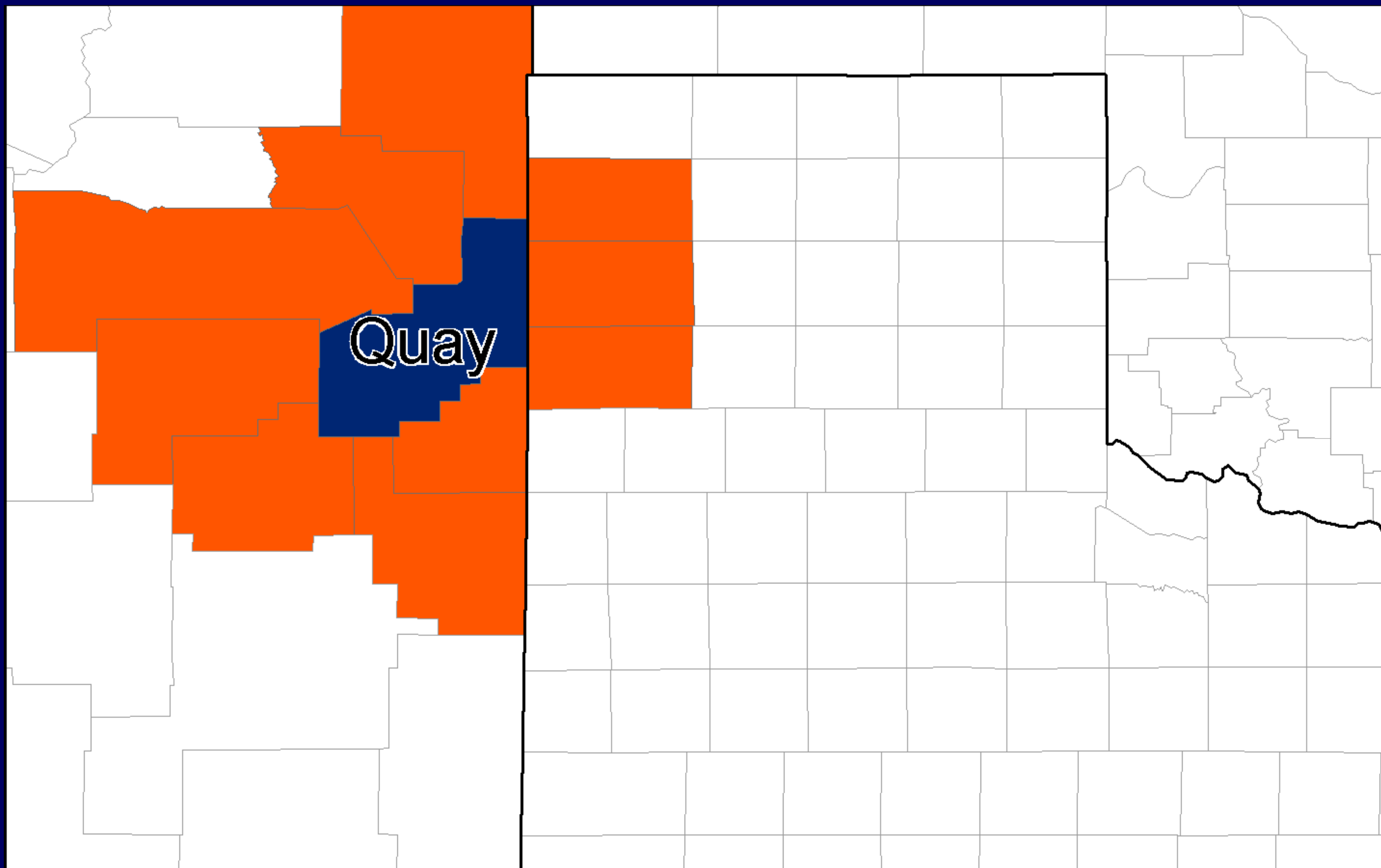
Real Life Problems

- SWBB may be biased if there is heterogeneity across space
  - Most common spatial weights matrices will oversample some regions
    - e.g., Edges of grid are undersampled
  - If variables are heterogeneous across space, estimates are averages
    - If regions are oversampled, they bias estimates of averages

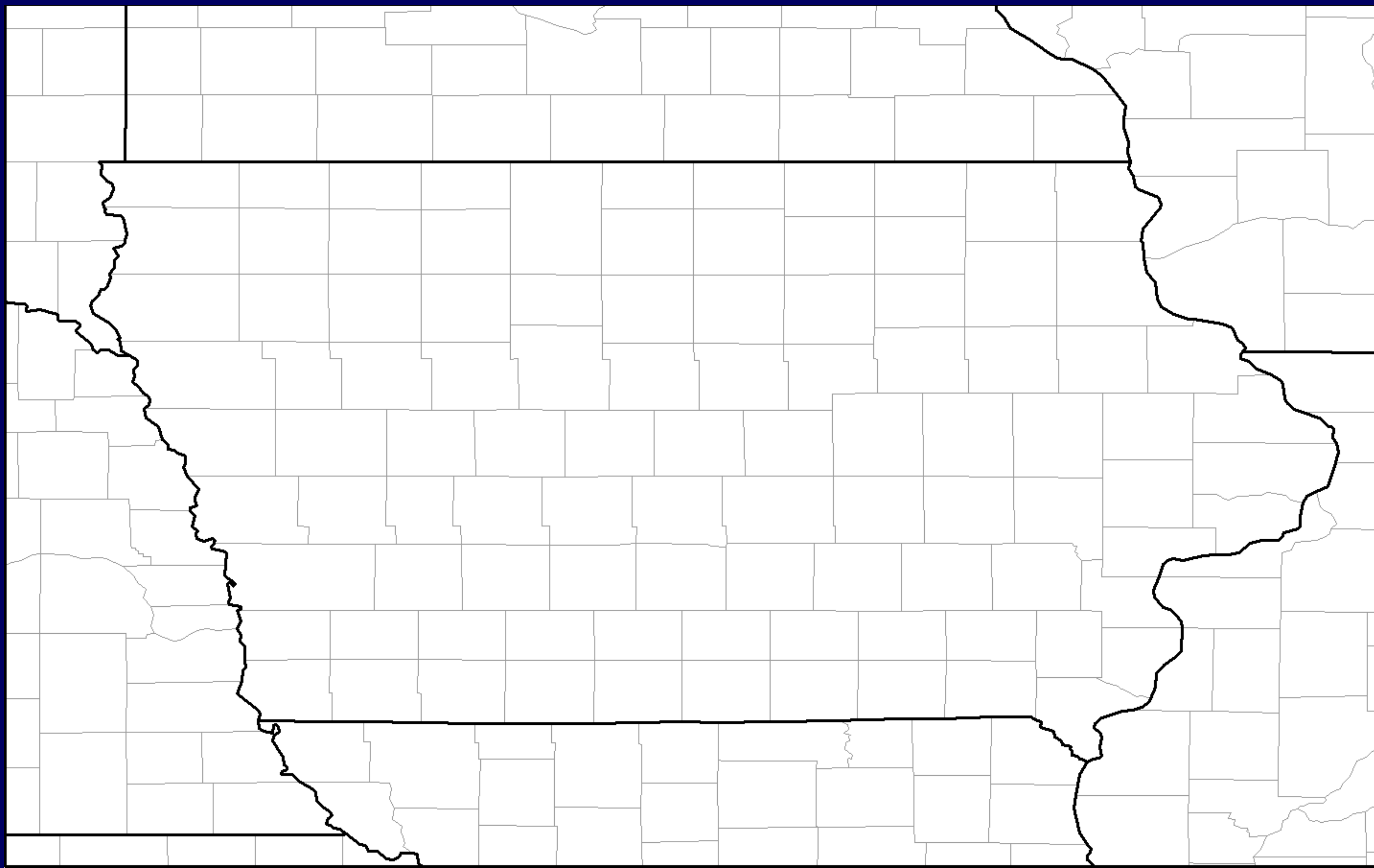
Real Life Problems



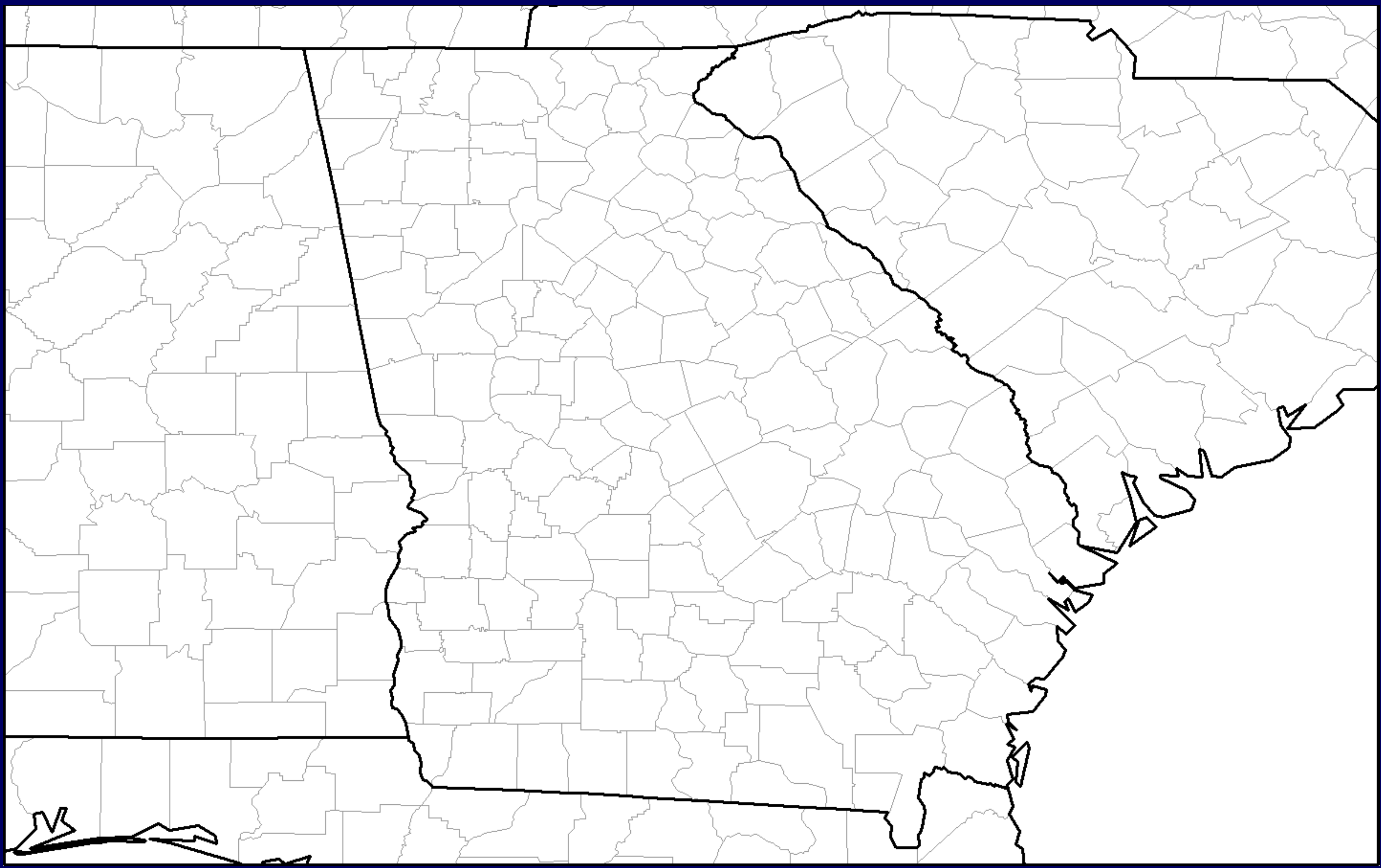
Real Life Problems



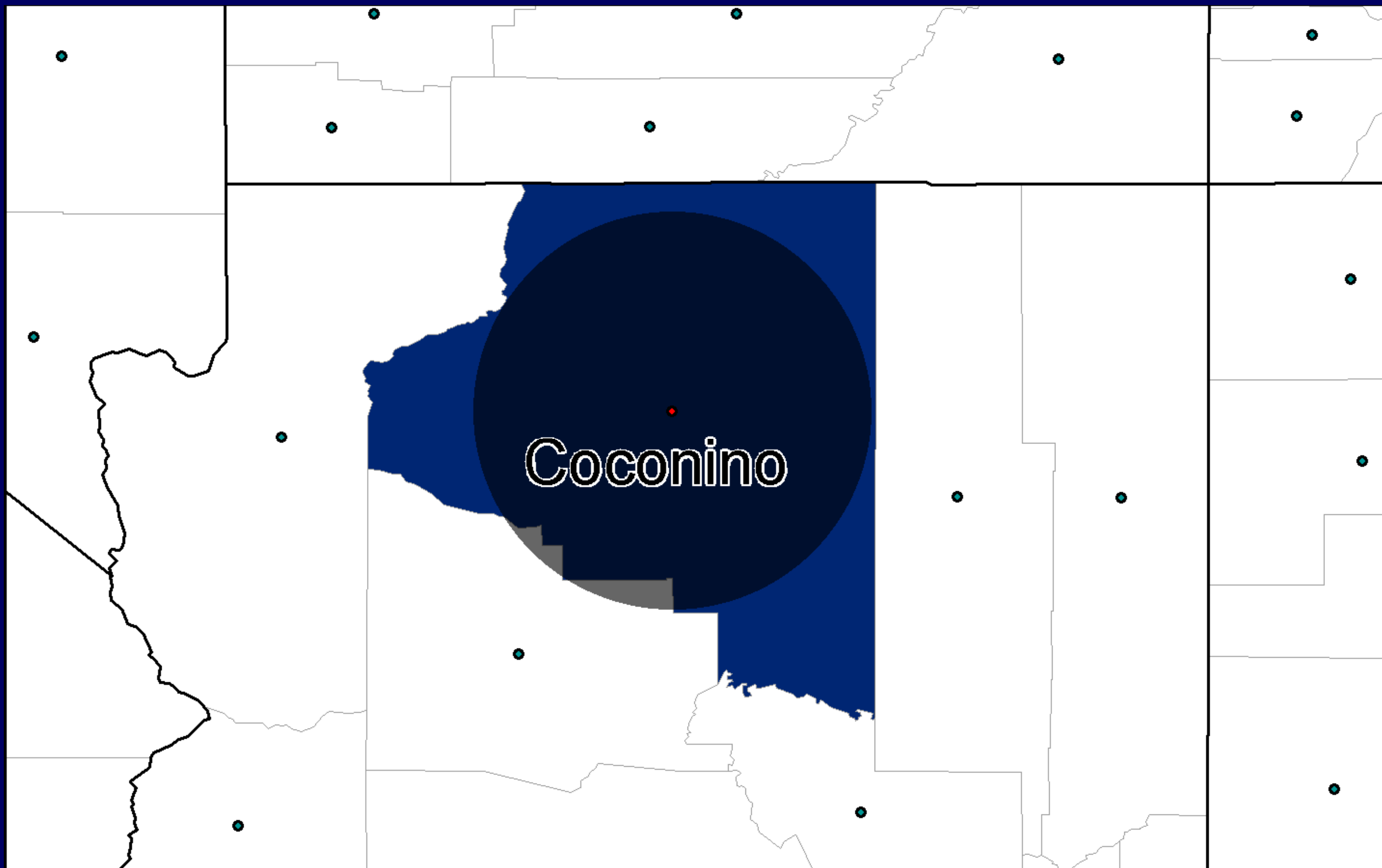
Real Life Problems



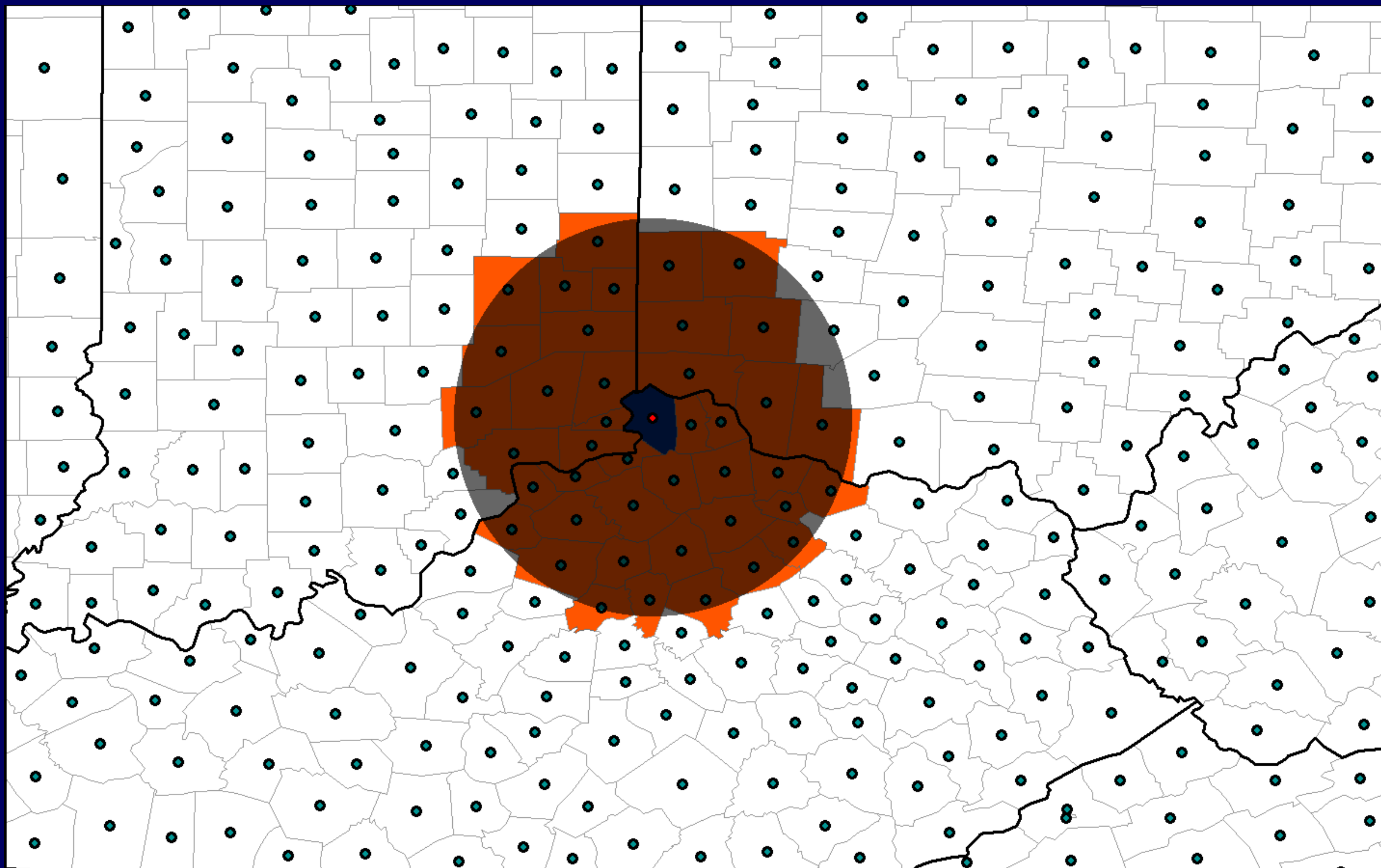
Real Life Problems



Real Life Problems



Real Life Problems



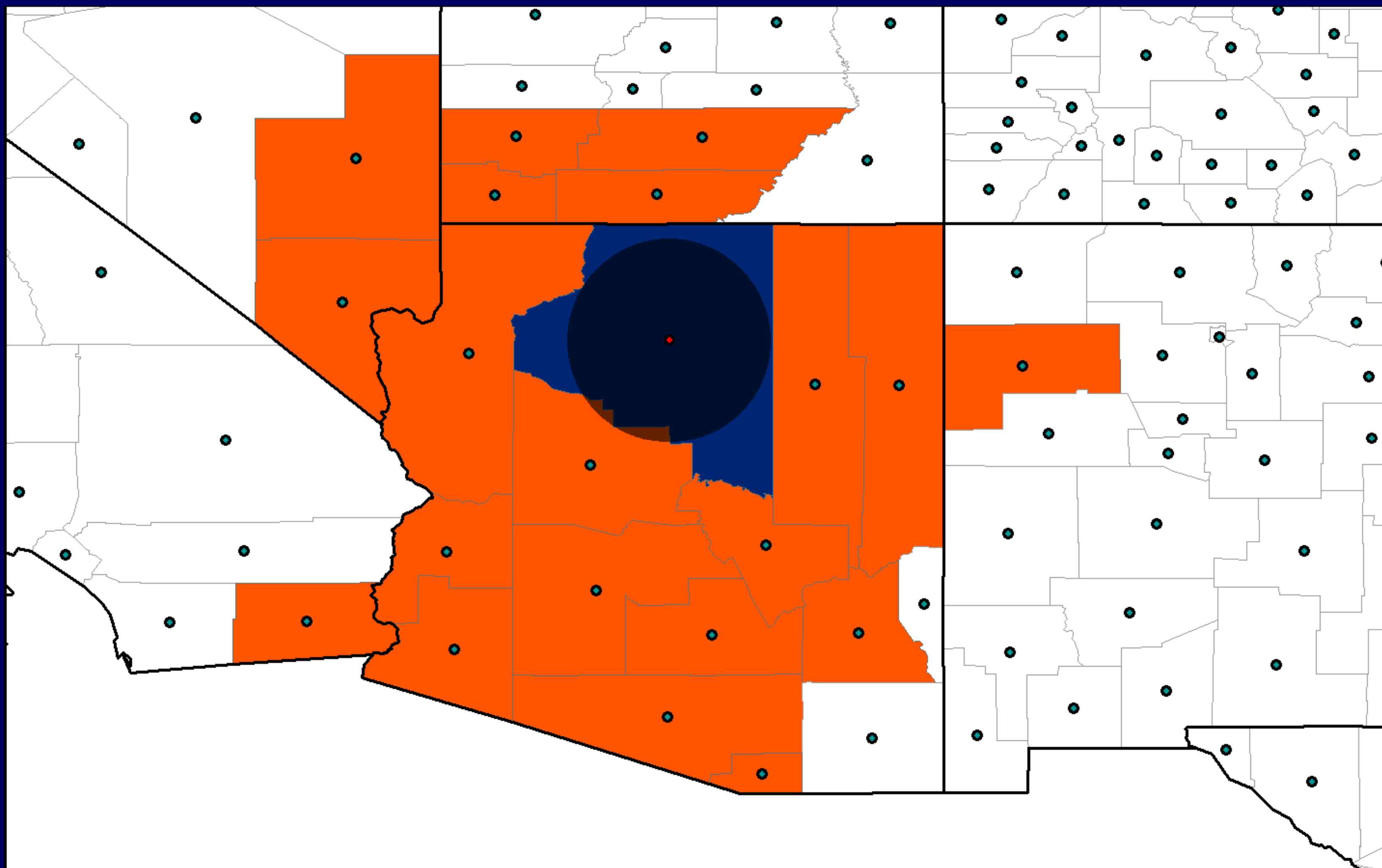
Real Life Problems

- Solution: Transposed Nearest k Neighbors Weights Matrix
- Nearest k Neighbors:
  - Which k neighbors are closest to me
- Transposed Nearest k Neighbors:
  - Which neighbors have me as one of their k nearest neighbors

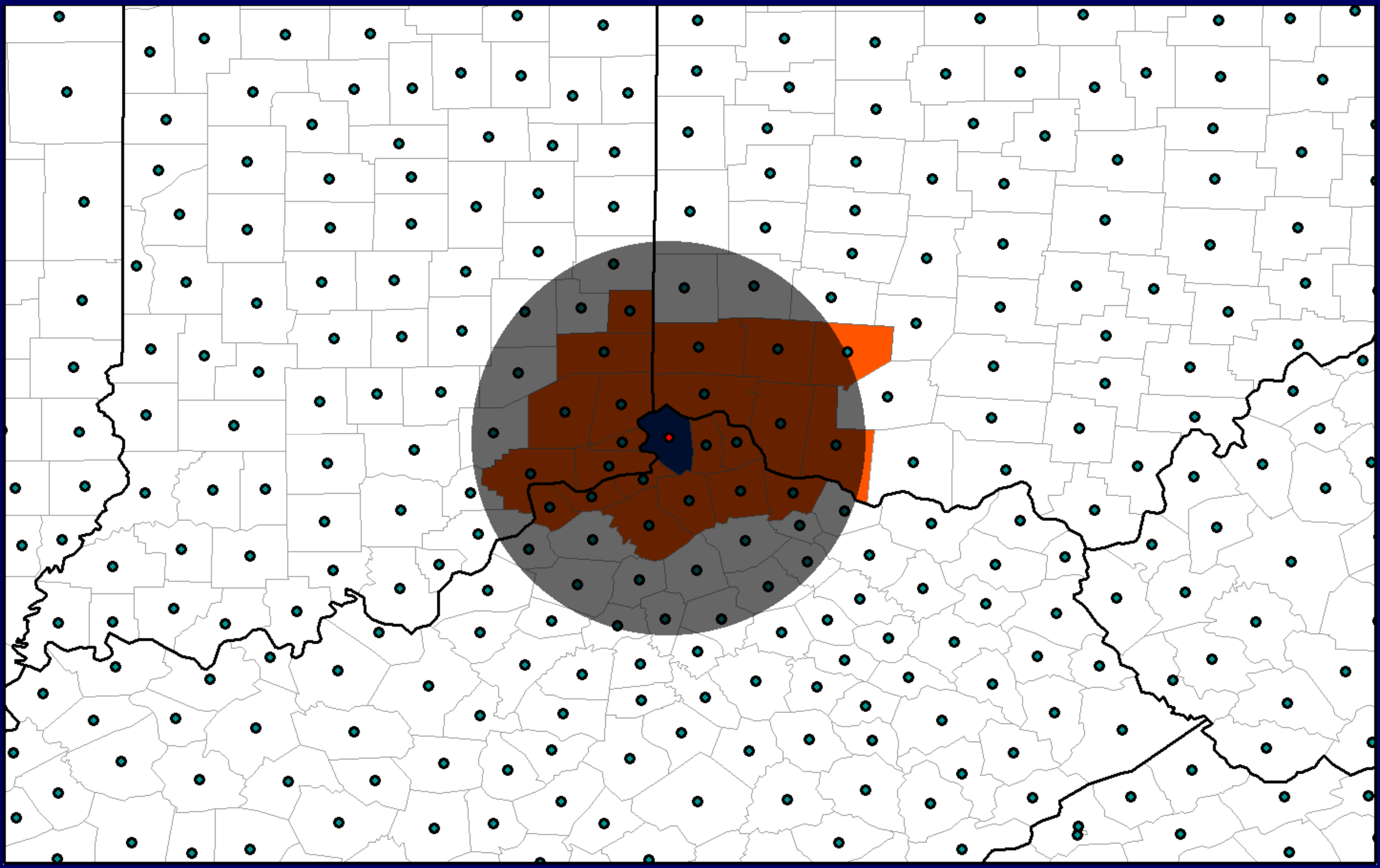
Real Life Problems

- Why Transposed Nearest k Neighbors Works:
  - Every county has equal chance of being drawn as a seed
  - Every county has k nearest neighbors
    - i.e. k in 3,048 chance of one of its nearest neighbors being drawn as seed
  - Thus every county has a k+1 chance in 3,048 of being drawn in any draw of a seed and its neighbors

Real Life Problems



Real Life Problems



Drake Warren
dewarren@uiuc.edu