AREAL DATA

STA465: Theory and Methods for Complex Spatial Data

Instructor: Dr. Vianey Leos Barajas

ANNOUNCEMENTS

- Hwk 1 Solution is online

- Hwk 2 is posted and due on Feb 22nd by 11:59 pm EST

- Reading week: February 15-19 (no class)

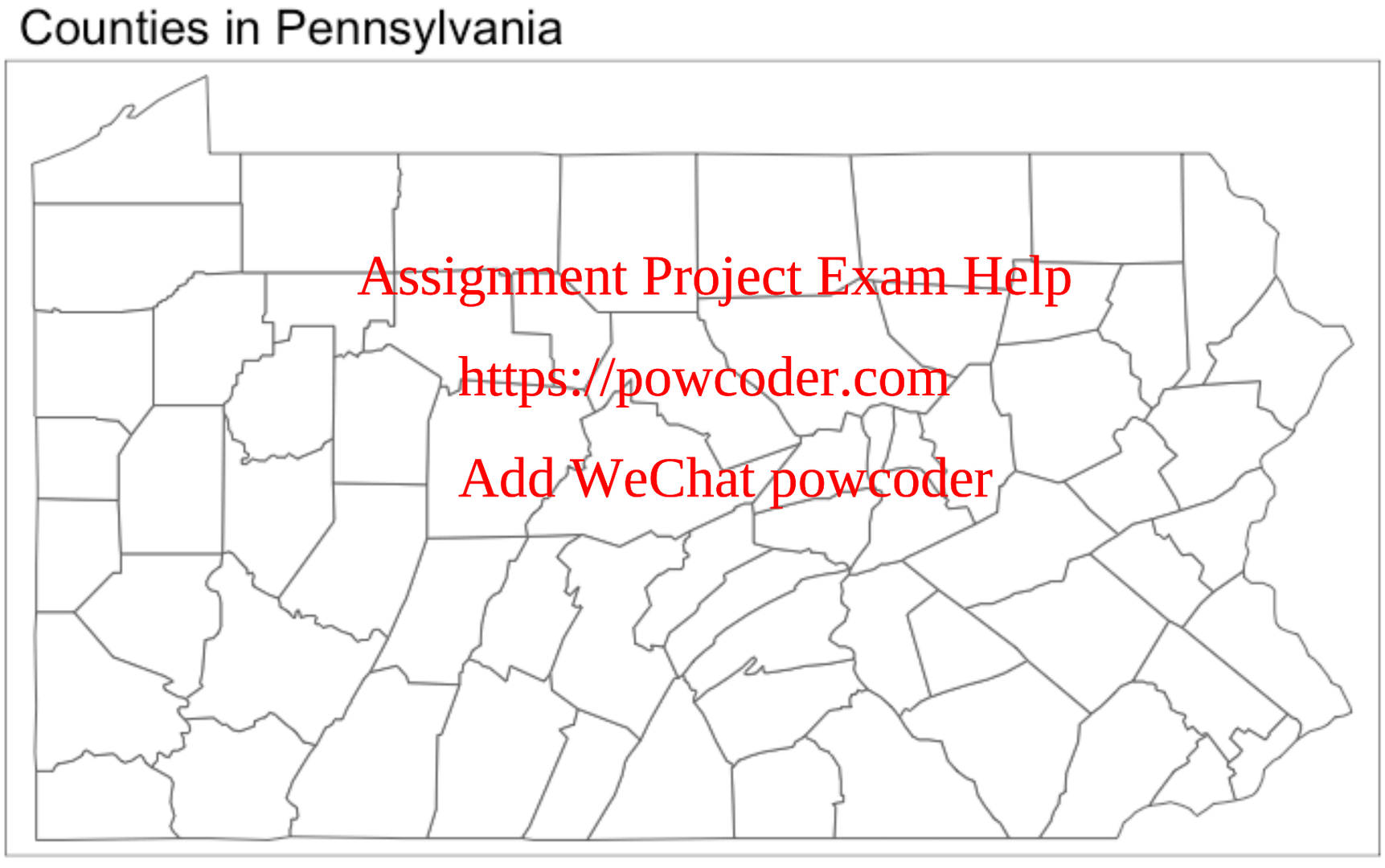
WHAT IS AREAL DATA?

- **Areal or lattice data — when a fixed domain is partitioned into a finite number of subregions.**

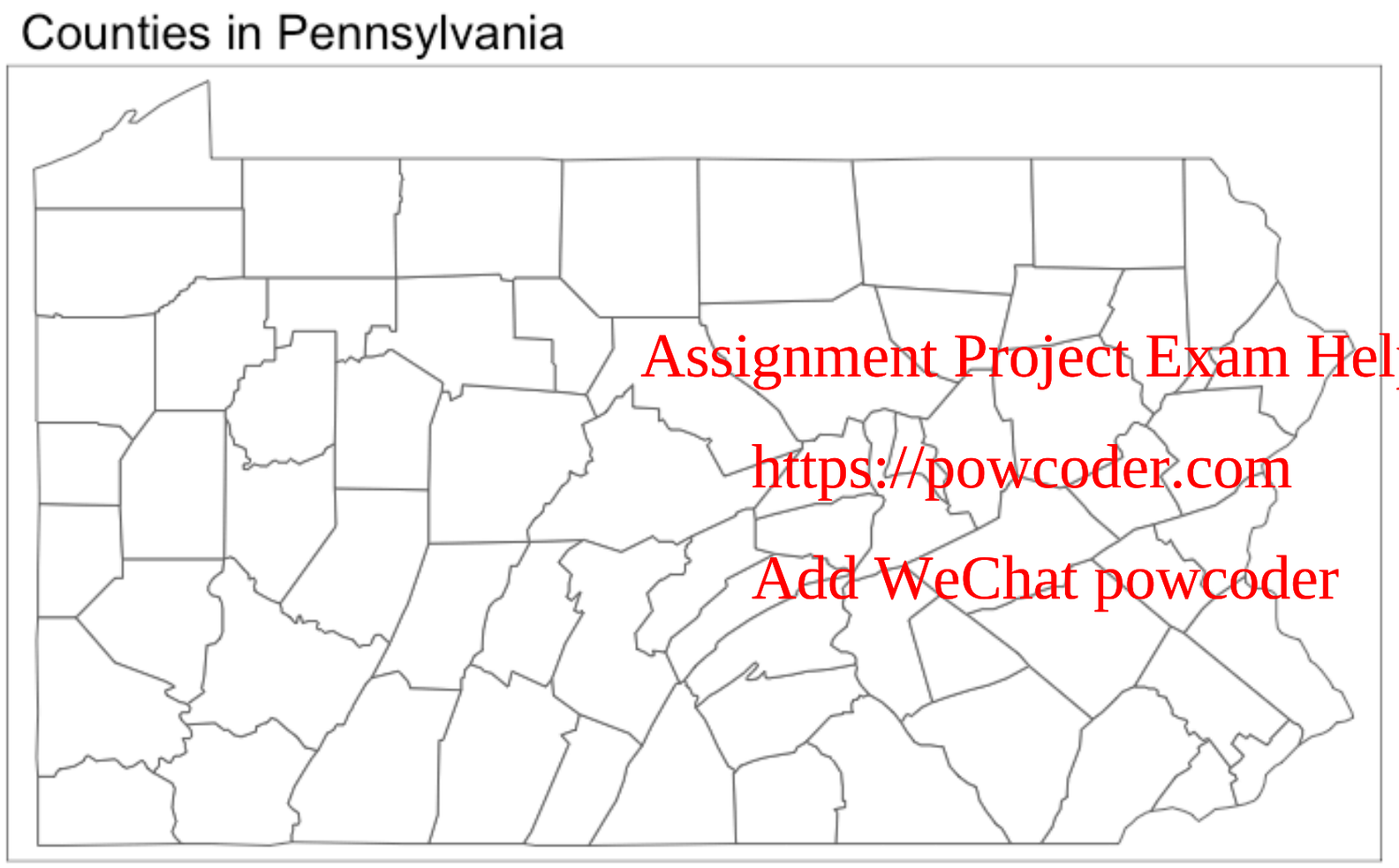
- Defined on a finite or countable subset in space — e.g. grid nodes, pixels, polygons, small areas

- Examples:
  - Number of cancer cases in counties (USA)
  - Number of road accidents in provinces
  - Proportion of people living in poverty in census tracts

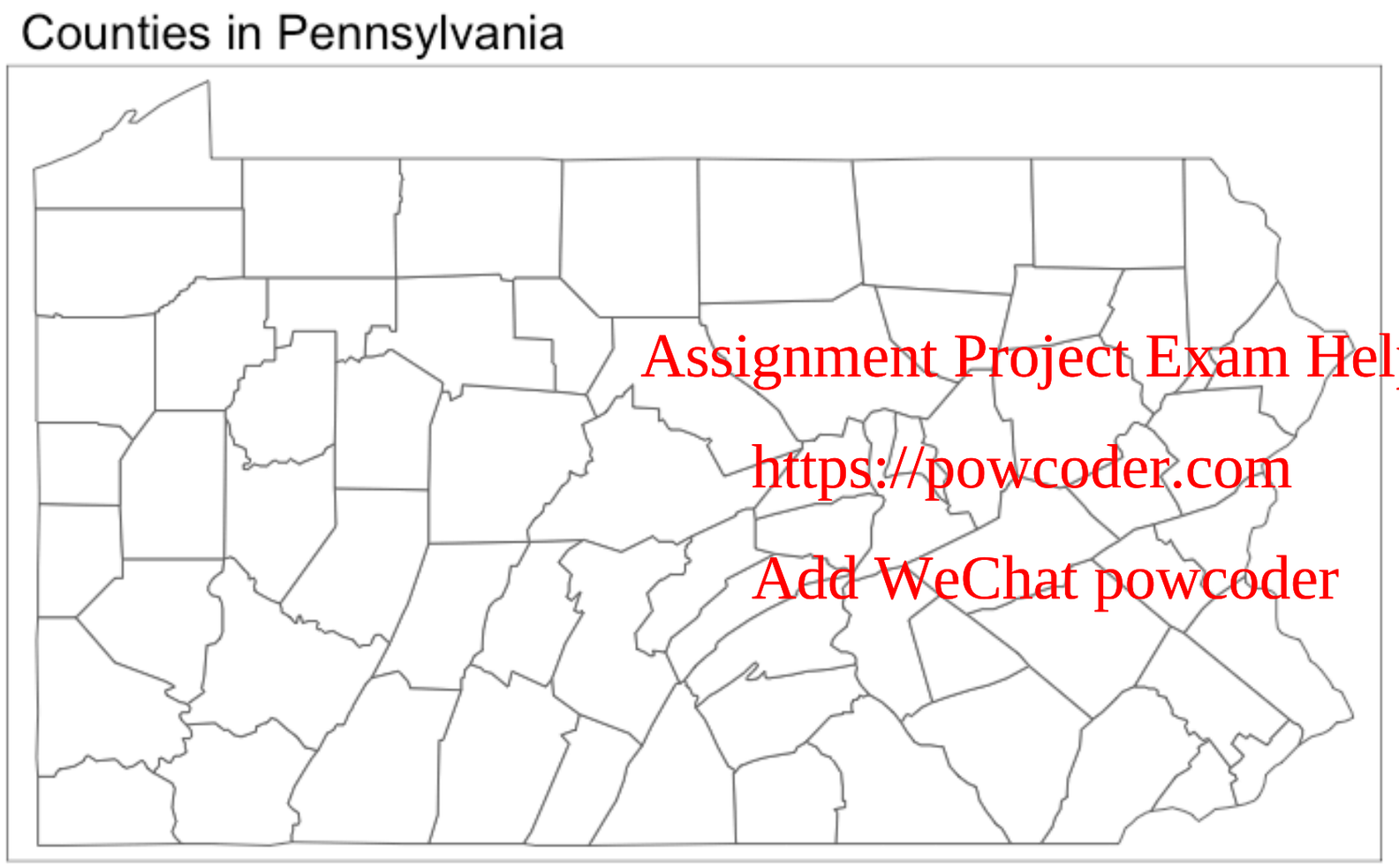
EXAMPLE DATA



SPATIAL NEIGHBOURS — ADJACENCY



SPATIAL NEIGHBOURS — DISTANCE



SPATIAL NEIGHBORHOOD MATRIX

- Spatial neighborhood matrix: W
- $(i,j)th$ element of spatial neighborhood denoted by $w_{i,j}$ spatially connects areas i and j
  - Elements can be viewed as 'weights'
  - More weight is associated with j's closer to i than those farther away from I

SPATIAL NEIGHBORHOOD MATRIX

- Simplest neighborhood definition:
  - $w_{ij} = 1$ if regions i and j share common boundary
  - $w_{ij} = 0$ otherwise

- Customarily, $w_{ii} = 0$

ELEMENTS OF A SPATIAL MODEL

BESAG-YORK-MOLLIÉ MODEL

- Takes into account that data may be spatially correlated

- Observations in neighbouring areas may be more similar than observations in areas that are farther away

- Model includes:
  - Spatial random effect that smoothes the data according to a neighborhood structure
  - Unstructured exchangeable component that models uncorrelated noise

BESAG-YORK-MOLLIÉ MODEL: EXAMPLE

- Assume we are interested in observed counts Y_i in area i for spatial small area disease risk estimation
- $Y_i \sim Po(E_i\theta_i), \quad i = 1,\ldots,n$
- $log(\theta_i) = \alpha + u_i + v_i$
- $v_i \sim N(0,\sigma_v^2)$

- E_i — expected counts
- θ_i — relative risk in area i

BESAG-YORK-MOLLIÉ MODEL: CAR

- $u_i \,|\, \boldsymbol{u}_{-i} \sim N\left(\bar{u}_{\delta_i}, \frac{\sigma_u^2}{n_{\delta_i}}\right)$
- $\bar{u}_{\delta_i} = n_{\delta_i}^{-1} \sum_{j \in \delta_i} u_j$
- $\delta_i =$ set of neighbours
- $n_{\delta_i} =$ number of neighbours over area i

NEXT TIME

- We'll put everything we've learned together:
  - Simulate from model with fixed values for parameters
  - Simulate from prior predictive
  - Fit the model
  - Simulation from posterior predictive
  - + Maps along the way!