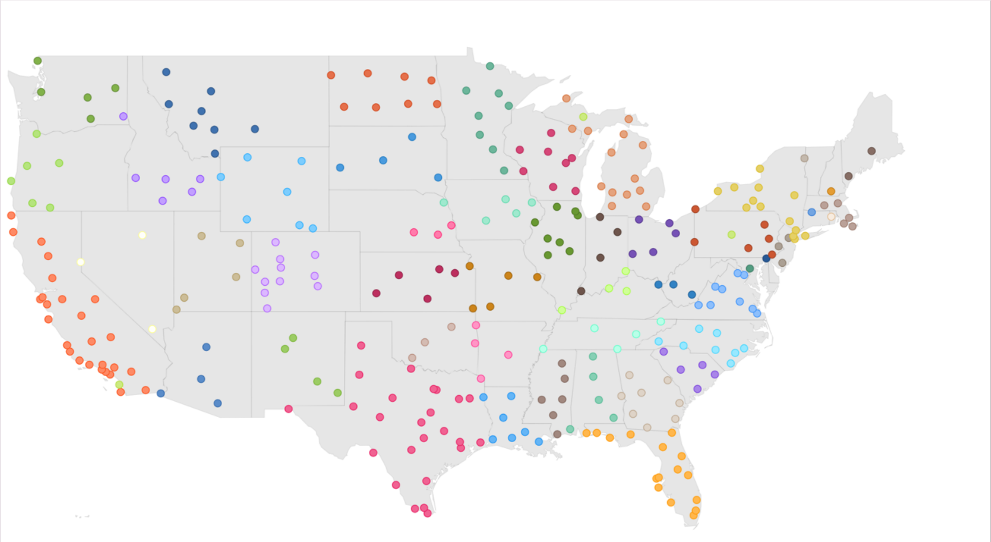
Bayesian Disease Mapping with R and OpenBUGS

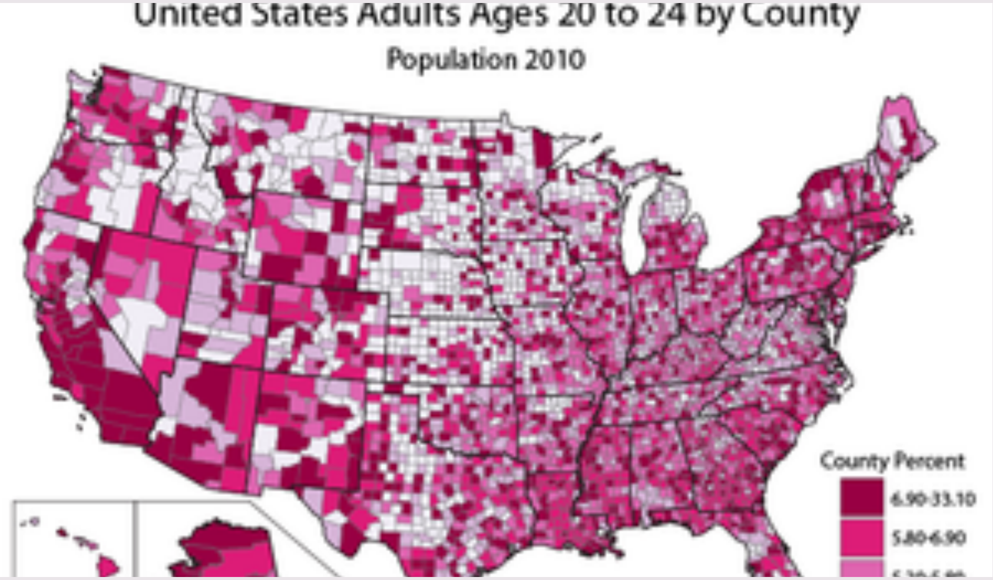
Seungwon Kim

Department of Geographical and Sustainability Science

Introduction

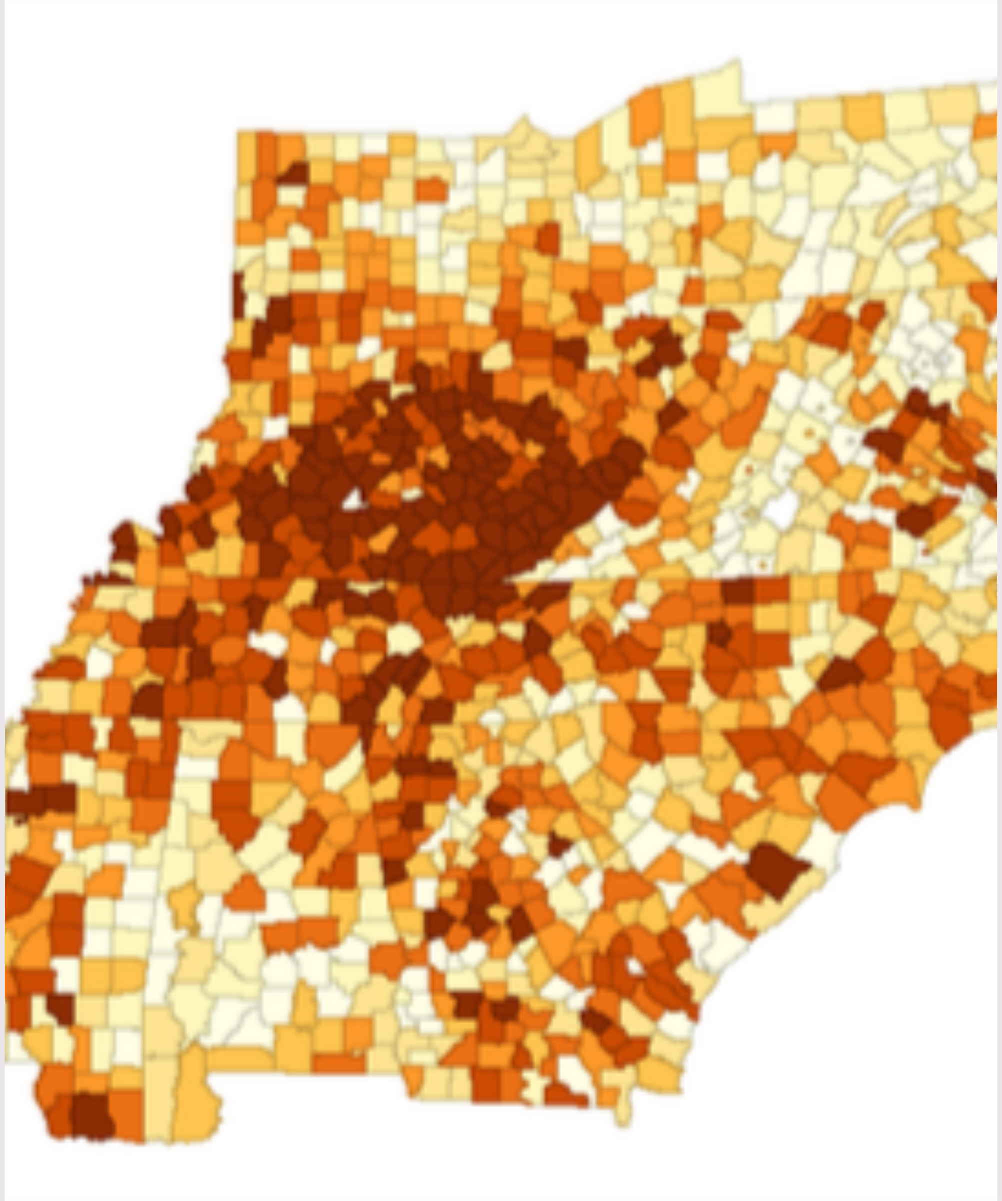
- Different ways to map health & crime related data
  - Scattermap
  - Choropleth map
    - Cases aggregated to geographic units (states, county, census track, etc..)
    - Generally mapped as rate
      - # of cases / # of population at risk





Introduction

- Average annual rates for lung cancer rates among males (2011 – 2015)
- The 65+ male population as the denominator
- Risk factors?
  - Smoking
  - Living in a heavy coal-mining area such as Kentucky – Exposed to pollution from mining activities



Issues with rates

- **Crude rate**
- $r_i = \frac{O_i}{P_i}$
  - O_i observed count at *i* and P_i the population at risk at *i*
- The use of crude rates to estimate rare disease risk in small areas is often problematic since these measures are subject to large variation
- Difficult to interpret & often misleading & unreliable rates

Issues with rates

- **Standard Mortality Rates (SMR) – Relative risk**
- $SMR_i = \frac{O_i}{E_i}, E_i = \frac{\Sigma O_i}{\Sigma P_i} \times P_i$
    - O_i observed count at *i* and E_i expected count at *i*
    - *High-risk area if* SMR_i *> 1*
    - *Average if* SMR_i*. = 1*
    - Low-risk area SMR_i *< 1*
    - O_i*=100,* E_i*=50 ->* $SMR_i = 2$
    - O_i*=3,* E_i*=1.5 ->* $SMR_i = 2$

Issues with rates

- Extreme ratios associated with areas with the smallest populations
- Bayesian disease mapping seeks to borrow information from neighboring areas to produce stable estimates.
- First law of geography
  - "Everything is related to everything else, but near things are more related than distant things"
- Improve the reliability of the observed rates

Source: https://www.slideshare.net/THLfi/eric-delmelle-disease-mapping

Bayesian inference

- Use Bayes' theorem to update the probability for a hypothesis as more information becomes available
- $P(H|E) = \frac{P(H)P(E|H)}{P(E)} \approx P(H)P(E|H)$

Poisson model

- Poisson distribution
  - A discrete probability distribution that express the probability of a given number of events occurring in a fixed interval of time or space
  - $P(Y = k|u) = \frac{u^k e^{-u}}{k!}$
  - $E(Y) = Var(Y) = u$

Poisson model

- Poisson regression
  - $y_i \sim Poisson(u_i)$
  - $E(Y|x) = e^{\boldsymbol{\theta}^T x}$
  - $\log(u_i) = \log(E_i) + \ a_i + b_1\, x_i$
  - $\log(\frac{u_i}{E_i}) = a_i + b_1\, x_i$
  - Relative risk $= \frac{u_i}{E_i} = e^{a_i + \ b_1\, x_i}$

Poisson model

- Spatial data
  - Not independent, but spatially dependent
  - Spatial autocorrelation
- Spatial model
  - $\log(u_i) = \log(E_i) + \ a_i + b_1 \, x_i + h_i$
  - h_i are spatial random effects assigned a CAR prior