Areal Simulation

Areal Data

- Represents data in a vectorized form within regular or irregular lattices
- Modeling Areal data can be used for disease mapping, ecological regression, other epidemiological use cases to name a few
- The goal of modeling Areal data is to understand a lattice's spatial auto correlation with the idea that polygons closer together would share more similarity in their measurement values than farther away

Spatial Auto Correlation

- Oftentimes it can be found that after adjusting for covariates some of the variance unexplained may be due to proximity; spatil correlation.
- Not accounting for spatial auto correlation violates the assumption of independence among the observations for regression models
- The effect of spatial auto correlation can be accounted for by modeling it's relationship with the outcome as a random effect and further augmenting the linear relationship between the outcome and covariates.

Conditional Autoregressive Models

Model Specification

- Spatial Units: Let S be a set of areal units k in n.
- Outcome: Let Y be the dependent variable with n values emerging from the same distribution
- Design Matrix: Let X be a set of p covariates with n observations
- Random effects: Let φ be a random effect for each k spatial unit
- Offset Term: Let O be a set of offset terms for each k spatial unit

Prior Specification 1

- For each β_k there is a normal distributional assumption with mean 0 and large variance
- General random effect prior in the multivariate case

$$\varphi \sim N(0, \tau^2 Q^{-1})$$

Where Q is a precision matrix that influence the spatial correlation for intrinsic models based on a weight matrix W that is a binary matrix of each $S_k, S_{j \neq k}$ according to the geographical contigueity leading to a Global CAR Prior

$$\varphi_k | \varphi_{-k} \sim N\left(\frac{\sum_i^n W_{ki}\varphi_i}{\sum_i^n W_{ki}}, \frac{\tau^2}{\sum_i^n W_{ki}}\right)$$

Average random effect with respect to its neighbors but may be overly smooth

Prior Specification 2

- Prior should account for varying levels of auto correlation; suggest by Lerox, et.al 1999

$$\varphi_k | \varphi_{-k} \sim N\left(\varrho \frac{\sum_i^n W_{ki} \varphi_i}{\varrho \sum_i^n W_{ki} + 1 - \varrho}, \frac{\tau^2}{\varrho \sum_i^n W_{ki} + 1 - \varrho}\right)$$

- Deemed to be the most effective until now

Prior Specification 3

- Local Priors will be covered here in more detail
- Note that these priors actually account for physical boundaries (lakes, rail roads, etc..) between area units measured as dissimilarities

Simulating Spatial Data

- Creating a weight matrix $\mathbf{W}$

```
     [,1] [,2] [,3] [,4]
[1,]    0    1    0    1
[2,]    1    0    0    1
[3,]    0    0    0    1
[4,]    1    1    1    0
```

- Creating a covariance matrix $\mathbf{C}$

```
     [,1] [,2] [,3] [,4]
[1,]  1.0  0.8  0.0  0.3
[2,]  0.8  1.0  0.0  0.3
[3,]  0.0  0.0  1.0  0.8
[4,]  0.3  0.3  0.8  1.0
```

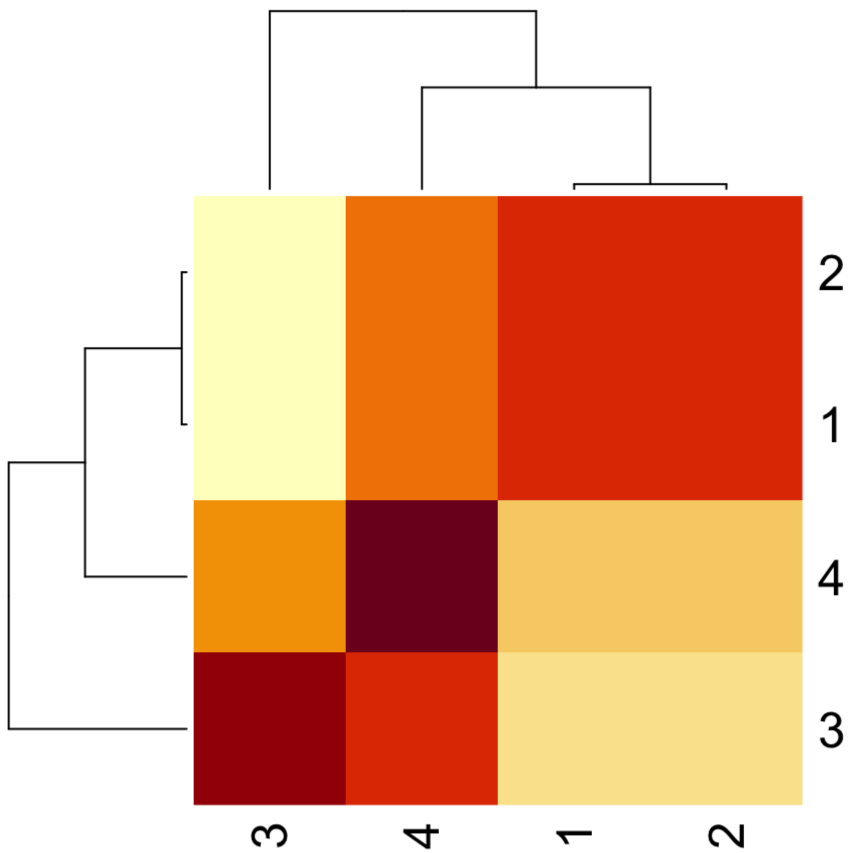
- Spatial data centered at 0, being 2 data points for each of the 4 areas.

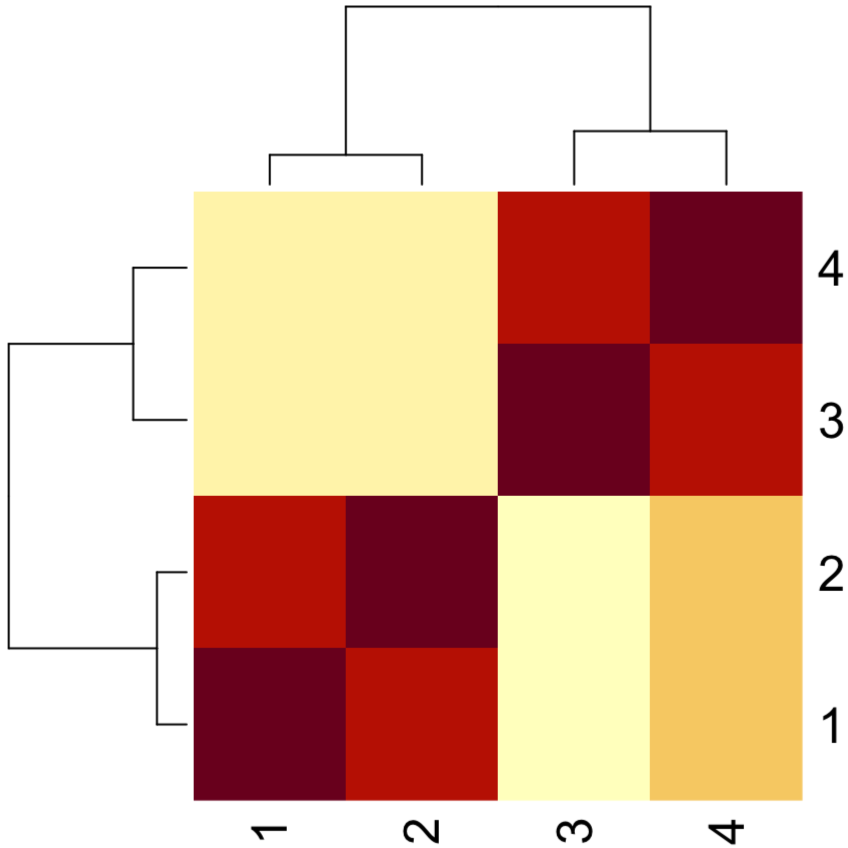
```r
R1 <- solve(diag(4) - 0.8 * W) %*% C %*% solve(diag(4) - 0.8 * t(W))
R2 <- solve(diag(4) - 0 * W) %*% C %*% solve(diag(4) - 0 * t(W))

sim_data1 <- mnormt::rmnorm(n = 2, mean = rep(0,4), varcov = R1)
sim_data2 <- mnormt::rmnorm(n = 2, mean = rep(0,4), varcov = R2)
```

```
         [,1]     [,2]       [,3]     [,4]
[1,] 2.712143 1.976846 -0.0551924 1.968759
[2,] 1.549700 1.651308  0.4152045 1.834268
```

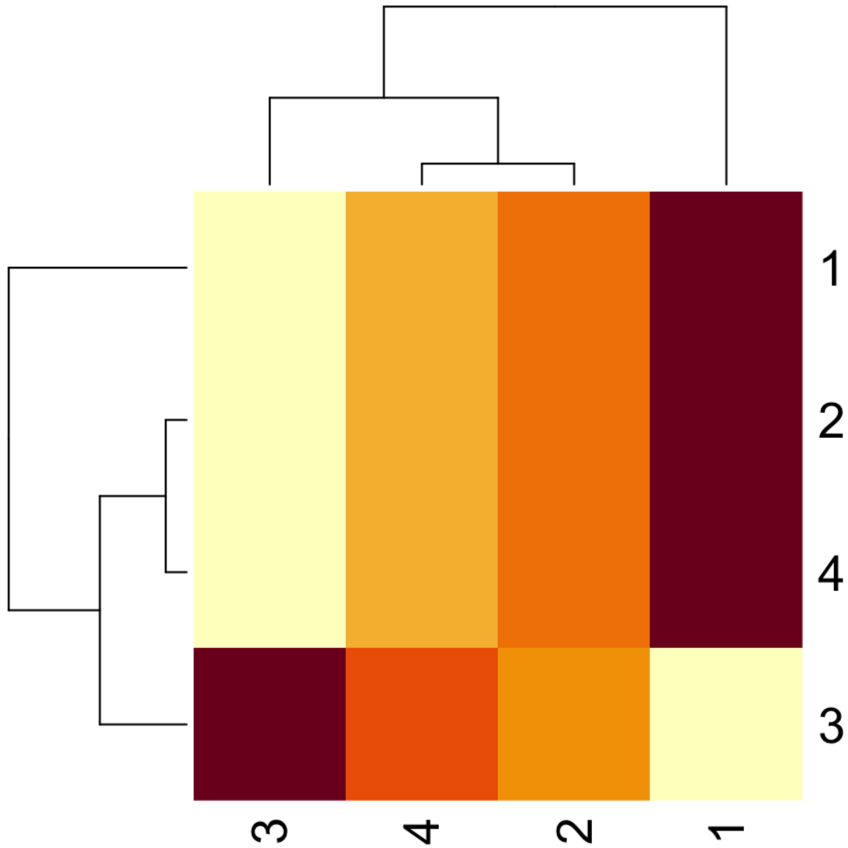
```
          [,1]        [,2]     [,3]      [,4]
[1,]  1.2764870 -0.01851471  1.306942  2.0609197
```

Recovering the spatial matrix





4
3
2
1
1
2
3
4


1
2
4
3
3
4
2
1