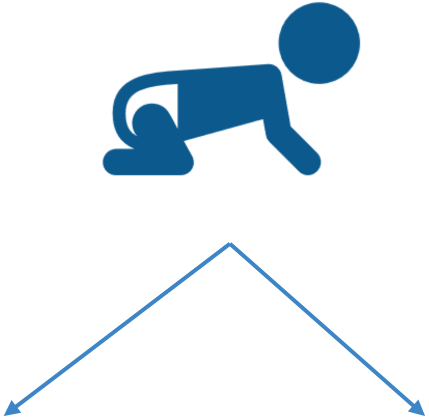
Modeling perinatal and neonatal mortality in Colombia and the U.S.

Mateo Rueda Molano
Shiva Ayoubzadeh

Motivation

Perinatal and Neonatal mortality

Perinatal:
Fetal period & early neonatal



Early neonatal:
Before 7 days of birth

Late neonatal:
7-28 days of birth

Motivation

Considerable differences in perinatal and neonatal mortality between developing and developed countries

Colombia: 32 states

$

U.S.: 50 states

$$$

Problem

This project will attempt to **model the probability of perinatal and late neonatal mortality in 2020** in different regions in **Colombia** and compare it with the **US**

Colombia: 32 states

$

U.S.: 50 states

$$$

Modeling idea

Deaths in the perinatal and neonatal period → **Treated as "successes"**

Births in the same year → **Treated as "trials"**

↓

Beta-binomial model:

The probability of success at each of n trials is not fixed, drawn from β distribution

Data

National Institute of Health (INS)

Epidemiological reports 2017-2020

No models found for this type of data

CDC wonder

Infant death records 2007-2020

Only predictive modeling for whole country

Models

Separate and hierarchical

Parameter of interest

$\theta \rightarrow$ Probability of perinatal and late neonatal death

Variables

y$\longrightarrow$ Number of deaths

n$\longrightarrow$ Number of births in the same period

Similarities

Likelihood

$$p(y_i|\theta_i) = \binom{n_i}{y_i} \theta_i^{y_i} (1 - \theta_i)^{y_i}$$

Both priors used death and birth means from the past data*

Priors and hyper-priors

Separate

$$\theta_i \sim Beta(\bar{a}_i, \bar{b}_i)$$

$\bar{a}_i$ → Mean of deaths in region i, from past data*

$\bar{b}_i$ → Mean of live births in region i, from past data*

Hierarchical

$$\theta i \sim \mathrm{Beta}(\alpha, \beta)$$

$$\alpha = \mu\eta, \beta = (1 - \mu)\eta$$

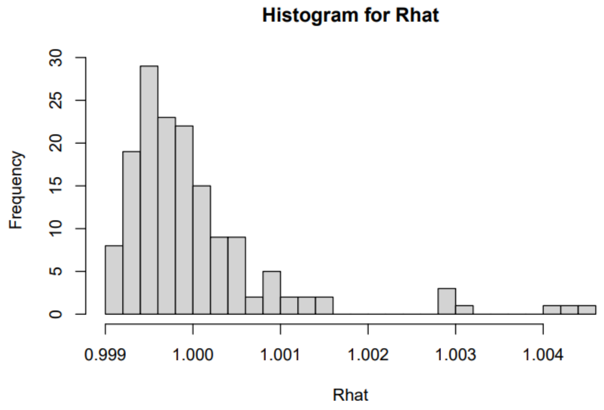
hyper-priors

$$\eta \sim \mathrm{exponential}(x)$$

$$\mu \sim Beta(\bar{\alpha}_i, \bar{\beta}_i)$$

Convergence

US Data



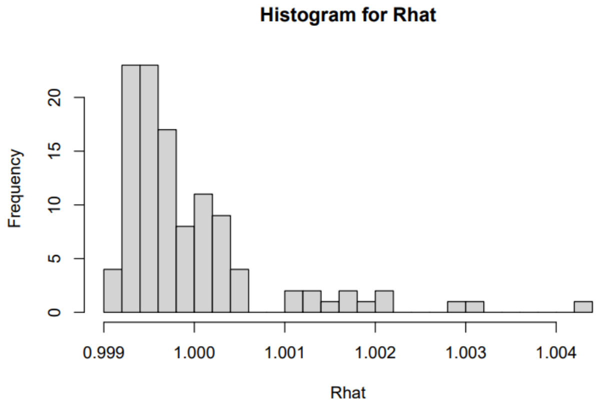

```
check_divergences(separate_modelUS)
```

```
## 0 of 4000 iterations ended with a divergence.
```

```
check_treedepth(separate_modelUS)
```

```
## 0 of 4000 iterations saturated the maximum tree depth of 10.
```

Colombia Data




```
check_divergences(separate_modelCol)
```

```
## 0 of 4000 iterations ended with a divergence.
```

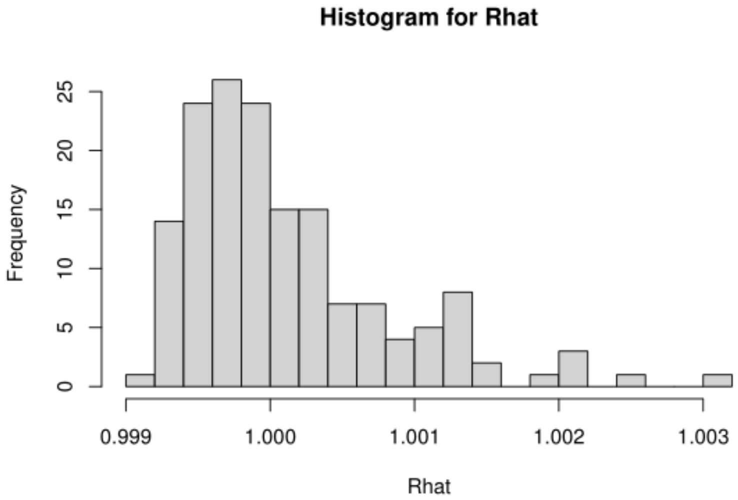
```
check_treedepth(separate_modelCol)
```

```
## 0 of 4000 iterations saturated the maximum tree depth of 10.
```

SEPARATE

Convergence

US Data



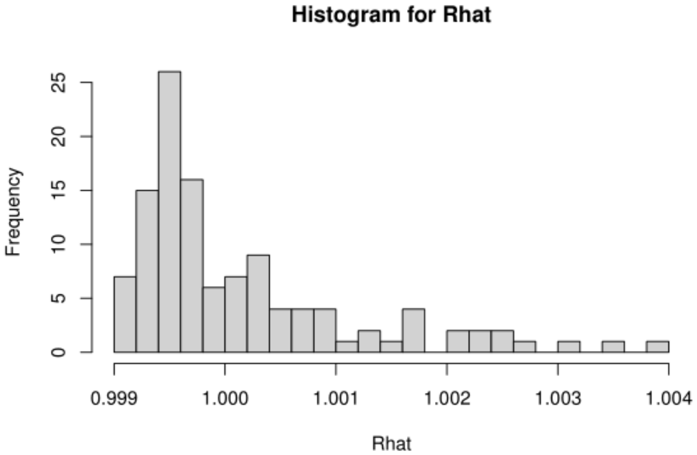

```
check_divergences(hierarchical_modelUS)
```

```
## 0 of 4000 iterations ended with a divergence.
```

```
check_treedepth(hierarchical_modelUS)
```

```
## 0 of 4000 iterations saturated the maximum tree depth of 10.
```

Colombia Data




```
check_divergences(hierarchical_modelCol)
```

```
## 0 of 4000 iterations ended with a divergence.
```

```
check_treedepth(hierarchical_modelCol)
```

```
## 0 of 4000 iterations saturated the maximum tree depth of 10.
```

HIERARCHICAL

Model improvements

Separate model

Hyper-priors

a[i] ~ normal(aMean[i], aStd[i])

b[i] ~ normal(bMean[i], bStd[i])

Prior

theta[j] ~ beta(a[j], b[j])

K-pareto analysis worse

Hierarchical model

$$\eta \sim \text{exponential}(x)$$

Instead of…

$$\eta \sim \text{gamma}(s,t)$$

Many s,t values were tried, k-pareto got worse

Model comparison

Separate model

US Data

```
##
## Computed from 4000 by 51 log-likelihood matrix
##
##           Estimate   SE
## elpd_loo    -233.2  9.6
## p_loo         31.4  4.7
## looic        466.4 19.2
## ------
## Monte Carlo SE of elpd_loo is NA.
##
## Pareto k diagnostic values:
##                          Count Pct.    Min. n_eff
## (-Inf, 0.5]   (good)     14    27.5%   703
##  (0.5, 0.7]   (ok)       25    49.0%   116
##    (0.7, 1]   (bad)      11    21.6%   31
##    (1, Inf)   (very bad)  1     2.0%   19
## See help('pareto-k-diagnostic') for details.
```

Colombia Data

```
##
## Computed from 4000 by 37 log-likelihood matrix
##
##           Estimate   SE
## elpd_loo    -164.6  6.7
## p_loo         21.4  3.8
## looic        329.1 13.4
## ------
## Monte Carlo SE of elpd_loo is NA.
##
## Pareto k diagnostic values:
##                          Count Pct.    Min. n_eff
## (-Inf, 0.5]   (good)      9    24.3%   366
##  (0.5, 0.7]   (ok)       22    59.5%   367
##    (0.7, 1]   (bad)       5    13.5%   48
##    (1, Inf)   (very bad)  1     2.7%   36
## See help('pareto-k-diagnostic') for details.
```

Hierarchical model

US Data

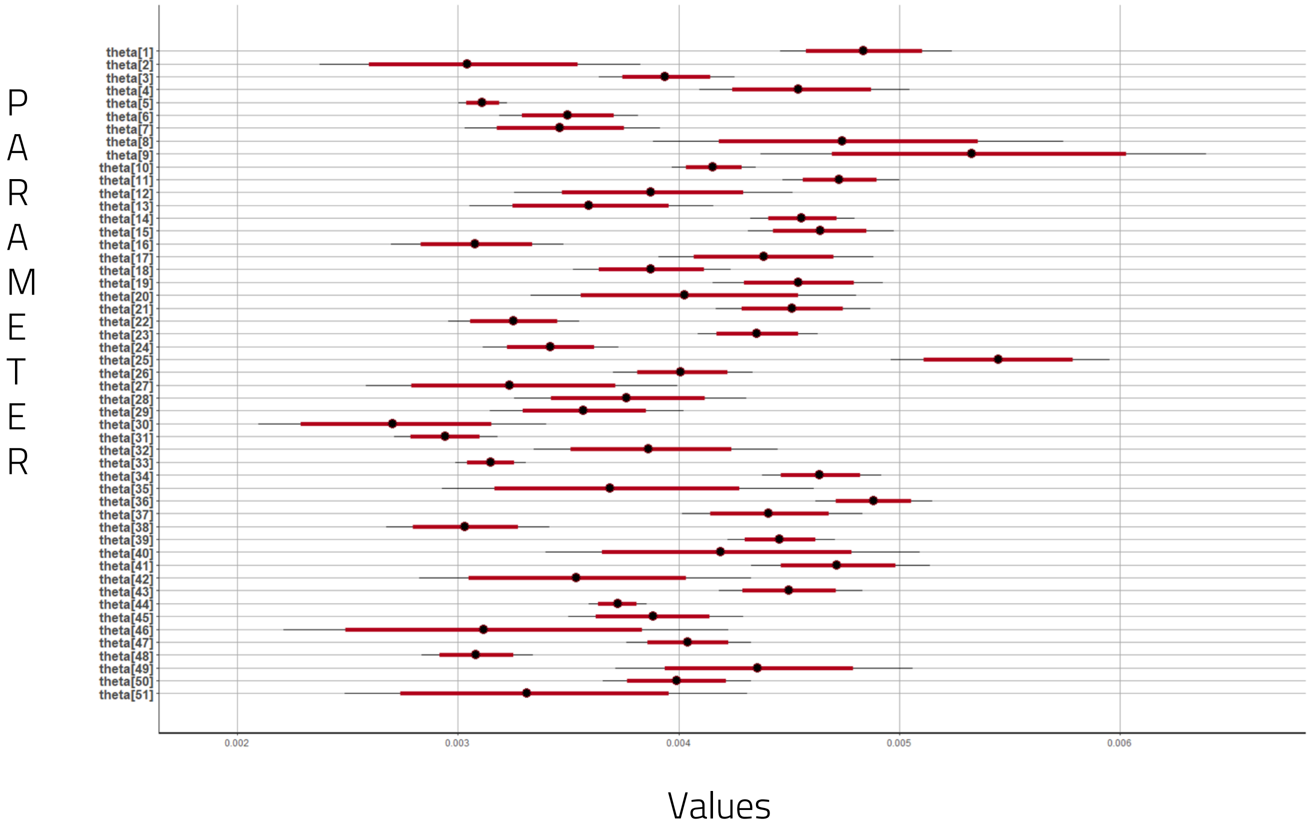
```
## Monte Carlo SE of elpd_loo is NA.
##
## Pareto k diagnostic values:
##                          Count Pct.    Min. n_eff
## (-Inf, 0.5]   (good)      3     5.9%   750
##  (0.5, 0.7]   (ok)       12    23.5%   230
##    (0.7, 1]   (bad)      30    58.8%   16
##    (1, Inf)   (very bad)  6    11.8%   15
## See help('pareto-k-diagnostic') for details.
```

Colombia Data

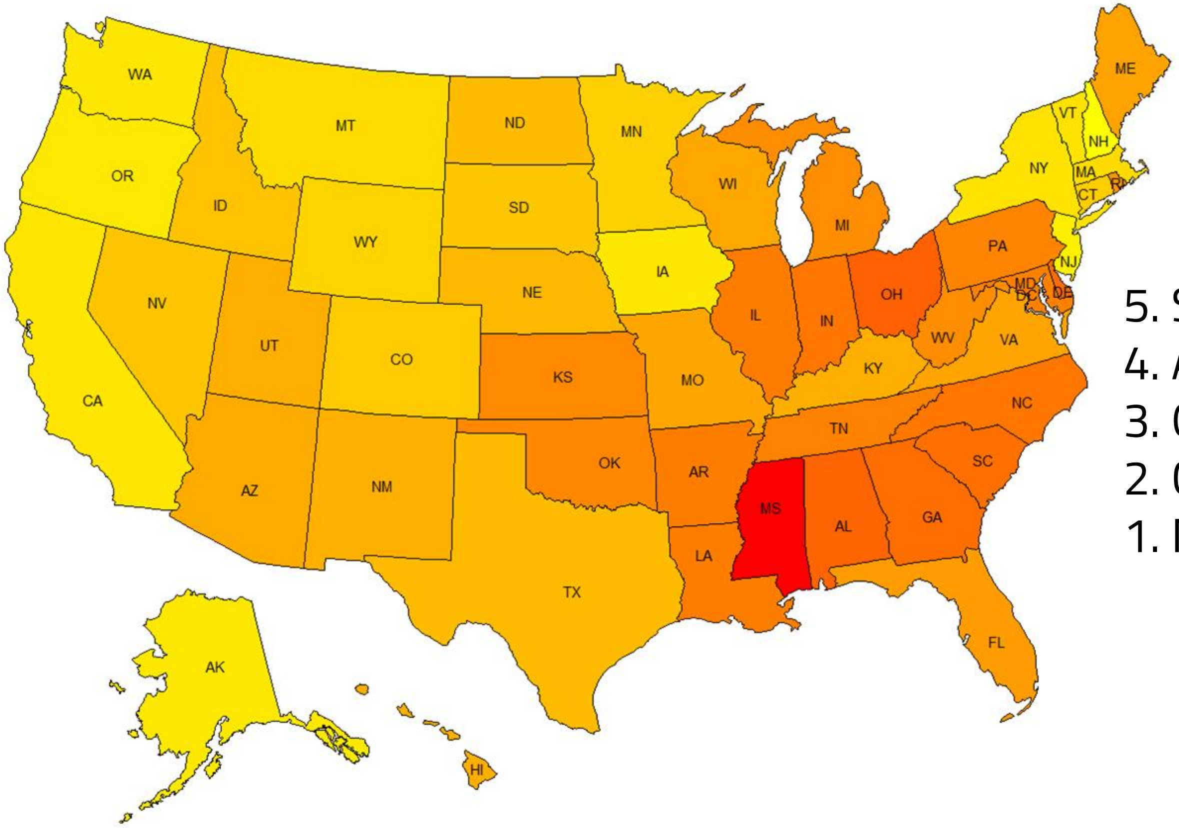
```
## Computed from 4000 by 37 log-likelihood matrix
##
##           Estimate   SE
## elpd_loo    -174.2  5.0
## p_loo         32.7  3.3
## looic        348.5 10.1
## ------
## Monte Carlo SE of elpd_loo is NA.
##
## Pareto k diagnostic values:
##                          Count Pct.    Min. n_eff
## (-Inf, 0.5]   (good)      0     0.0%   <NA>
##  (0.5, 0.7]   (ok)        6    16.2%   125
##    (0.7, 1]   (bad)      25    67.6%   21
##    (1, Inf)   (very bad)  6    16.2%   9
## See help('pareto-k-diagnostic') for details.
```

RESULTS

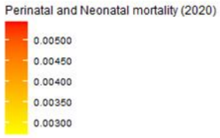
US DATA


P
A
R
A
M
E
T
E
R
theta[1]
theta[2]
theta[3]
theta[4]
theta[5]
theta[6]
theta[7]
theta[8]
theta[9]
theta[10]
theta[11]
theta[12]
theta[13]
theta[14]
theta[15]
theta[16]
theta[17]
theta[18]
theta[19]
theta[20]
theta[21]
theta[22]
theta[23]
theta[24]
theta[25]
theta[26]
theta[27]
theta[28]
theta[29]
theta[30]
theta[31]
theta[32]
theta[33]
theta[34]
theta[35]
theta[36]
theta[37]
theta[38]
theta[39]
theta[40]
theta[41]
theta[42]
theta[43]
theta[44]
theta[45]
theta[46]
theta[47]
theta[48]
theta[49]
theta[50]
theta[51]
0.002
0.003
0.004
0.005
0.006
Values

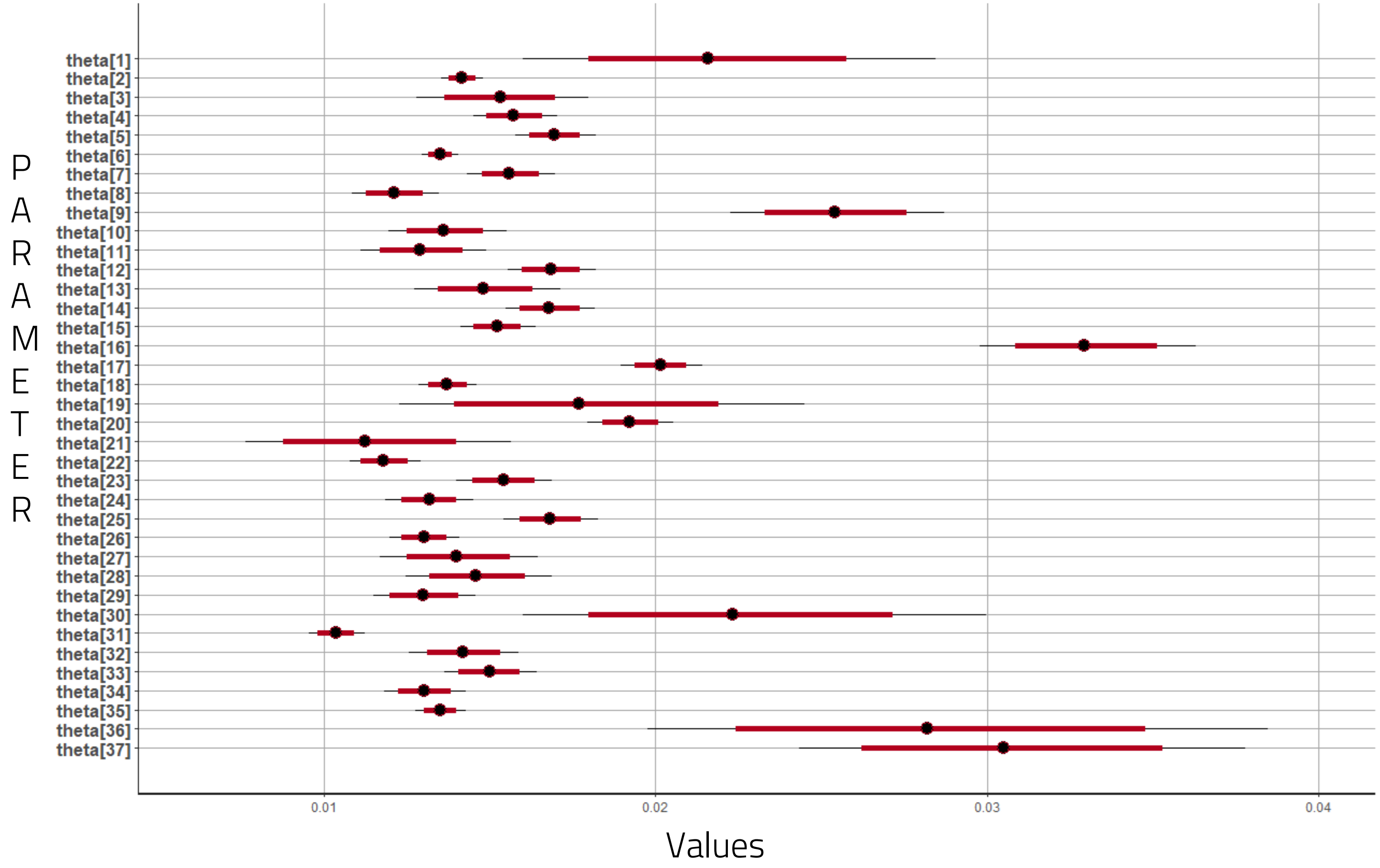
US

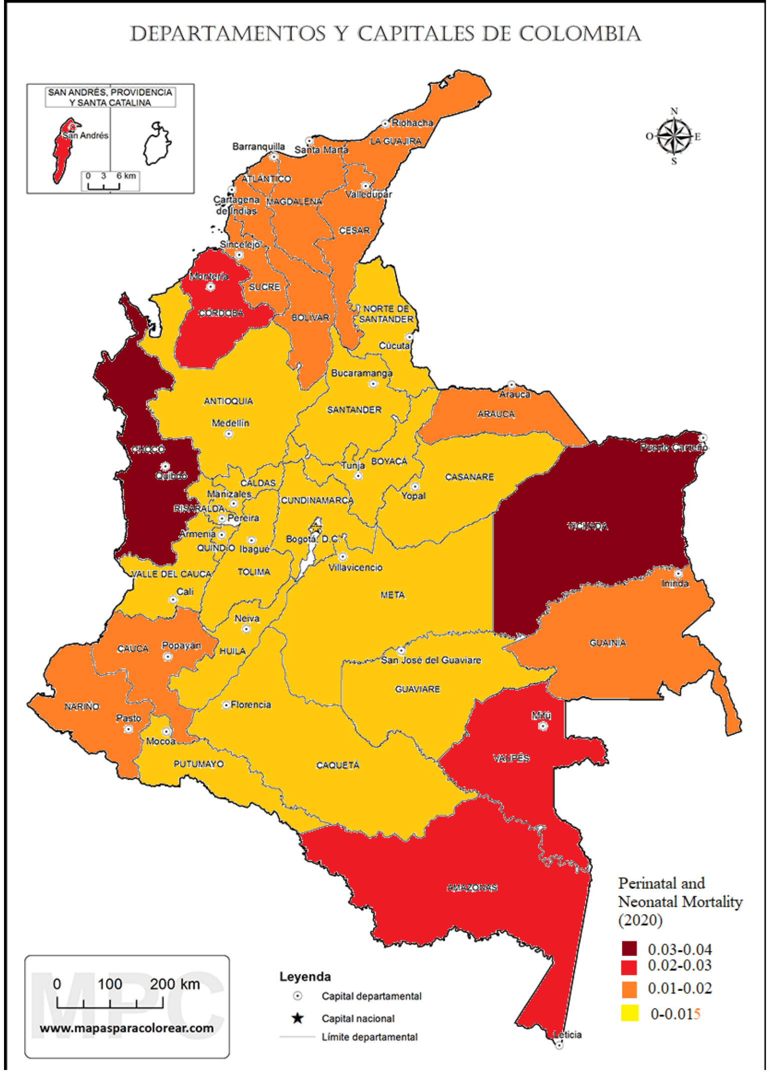


5. S. Carolina: 0.00472
4. Alabama: 0.00484
3. Ohio: 0.00488,
2. Columbia: 0.00534
1. Mississippi: 0.00545



COLOMBIA DATA





COLOMBIA

5. Cordoba: 0.0202
4. Buenaventura: 0.0284
3. Vaupes: 0.0284,
2. Vichada: 0.0306
1. Choco: 0.0330

Poorness vs Perinatal & Neonatal Mortality

Choco, Vichada and Vaupes are among the top 5 poorest states in Colombia [2]

South Carolina, Columbia, Mississippi and Alabama are between the top 10 poorest states in the US [3]

Discussion of Issues

Separate model

K-pareto analysis not optimal

Increase data size using not the mean but all past year data

Hierarchical model

K-pareto analysis not good

Try different hyper-priors and use all past data.

Interesting idea: Combine the model with a predictive one (using all data)→ Forecast probability of death in each state in the future years. → **Arima, ArimaX**

¿Any questions?

References

[1] “Neonatal Mortality.” UNICEF DATA, 20 July 2021, https://data.unicef.org/topic/child-survival/ neonatal-mortality/.

[2] Semana Magazine. “Pobreza En Colombia: Estos Son Los Departamentos Más Afectados.” Semana.com Últimas Noticias De Colombia y El Mundo, 21 Dec. 2020, https://www.semana.com/economia/articulo/pobreza-en-colombia-estos-son-los-departamentos-masafectados/202026/.

[3] November 10, 2021. “Top 10 Poorest States in the U.S.” Friends Committee On National Legislation, https://www.fcnl.org/updates/2021-11/top-10-poorest-states-us. Vehtari, Aki, and Markus Paasiniemi.

“BDA3 Demos Comparison of 2 Groups with Binomial.” Aalto University. Niemi, Jarad. “Hierarchical Models.” PPT file. Iowa State University, 2019. Hu, Jim Albert and Jingchen.

“Probability and Bayesian Modeling.” Chapter 10 Bayesian Hierarchical Modeling, 30 July 2020, https://bayesball.github.io/BOOK/bayesian-hierarchical-modeling.html. Brody-Moore, Peter,

“Bayesian Hierarchical Meta-Analysis of Asymptomatic Ebola Seroprevalence” (2019). CMC Senior Theses. 2228.https://scholarship.claremont.edu/cmc_theses/2228