FCSM 2018: Mapping Geographic and Temporal Variations in select Natality and Mortality Outcomes with R-INLA in Small Areas

Diba Khan
Centers for Disease Control and Prevention
National Center for Health Statistics
Division of Research and Methodology
Hyattsville, MD.

Disclaimer

The findings and conclusions in this presentation are those of the authors and do not necessarily represent the official positions of the National Center for Health Statistics, Centers for Disease Control and Prevention.

Co-authors and acknowledgement

1. Teen birth rates and visualization: Brady Hamilton, Lauren Rossen, Rong Wei, Yulei He, and Yinong Chong
2. Suicides mortality: Lauren Rossen, Holly Hedegaard, and Margaret Warner
3. Acknowledgement: Makram Talih and Jessica Keralis

Objective

To explore geographic and temporal variation in natality (teen births) and mortality (suicides) rates at the county level in U.S. using the National Vital Statistics data for the years 2003-2015 and 2005-2015 respectively.

Outline

1. **Bayesian philosophy**
2. **Integrated Nested Laplace Approximation**
3. **Example: Teen birth rates**
4. **Space time models**
5. **Model check and accuracy**
6. **Results**
7. **Data visualization**
8. **Example: Suicides mortality**
9. **Space time models**
10. **Model check and accuracy**
11. **Results**
12. **Conclusions**

Hierarchical Bayesian Model

1. Data Likelihood: $L(\mathbf{y}|parameters)$, where $\mathbf{y} = (y_1, y_2, ..., y_m)$, i=1,...,m areas
2. Prior distribution for the model parameters: Prob(parameters)
3. Prior distributions: uninformative or vague priors (are not assumption free)
4. Jeffreys priors are invariant to linear transformations but are improper
5. **Posterior distribution**: $\mathbf{Prob(parameters|y)} \propto \mathbf{L(y|parameters)} \cdot \mathbf{Prob(parameters)}$
6. Hierarchical Bayes: an extra level of hierarchy in setting the prior distribution of model parameters
7. Random effects: extra variability due to unmeasured confounders modeled by assigning an individual units effect

Flat priors

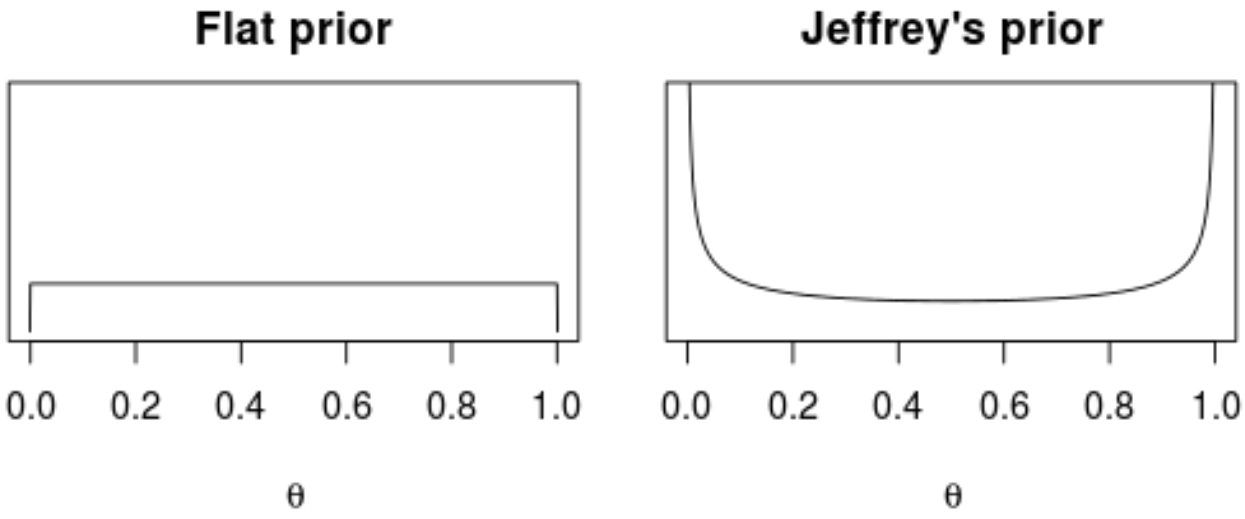


Figure: Priors.

Spatially correlated random effects

1. Spatially correlated effects: Intrinsic Conditional Autoregressive (ICAR) prior distribution
2. Spatial locations for areal data
3. Esri shapefile - polygon file gives geographical coordinates of the boundaries of each area
4. Weights are used to express spatial dependence between areas
5. Most commonly used specifications of weights: binary specification
6. The conditional expectation of ICAR prior random effect for an area is the average of the random effects in neighboring areas
7. The conditional variance of ICAR prior random effect for an area is inversely proportional to the number of neighbors

Posterior distribution: INLA, Winbugs, STAN, JAGS

1. Traditionally posterior distribution is estimated via Markov Chain Monte Carlo (MCMC): exact inference, extremely flexible - applicable to any type of data and model, **computational and time intensive**
2. **MCMC: Fundamental issues - model complexity and database dimension**
3. Gibbs sampling and Metropolis-Hastings available in WinBugs, OpenBUGS, JAGS
4. Posterior distribution can also be approximated via Laplace approximation in R-Integrated Nested Laplace Approximation (INLA) package: **computationally efficient alternative to MCMC**, **reliable estimates in less time**, **particularly relevant to large datasets**
5. Less established as compared to MCMC

R-INLA

1. Approximates the posterior marginals of a variety of Bayesian hierarchical models
2. Linear Mixed Models or Generalized Linear Mixed Models, Spatial, and Spatio-Temporal models, and point process or Geostatistical models
3. Approximation via Laplace integral approximation to the fixed effects
4. Numerical integration approximation to the random effects
5. **Exceedance probability: allows faster computation of the posterior probability that a parameter does/does not exceed a certain threshold**
6. More details in Havard Rue et al. (2009). JRSS-Series B. *Approximate Bayesian Inference for latent Gaussian models by using integrated nested Laplace approximations.*

Small area direct estimates: example of county level teen births for the age group: 15-19 for the years 2003-2015

Data suppression: Births less than 20 suppressed, 15-19, 2015

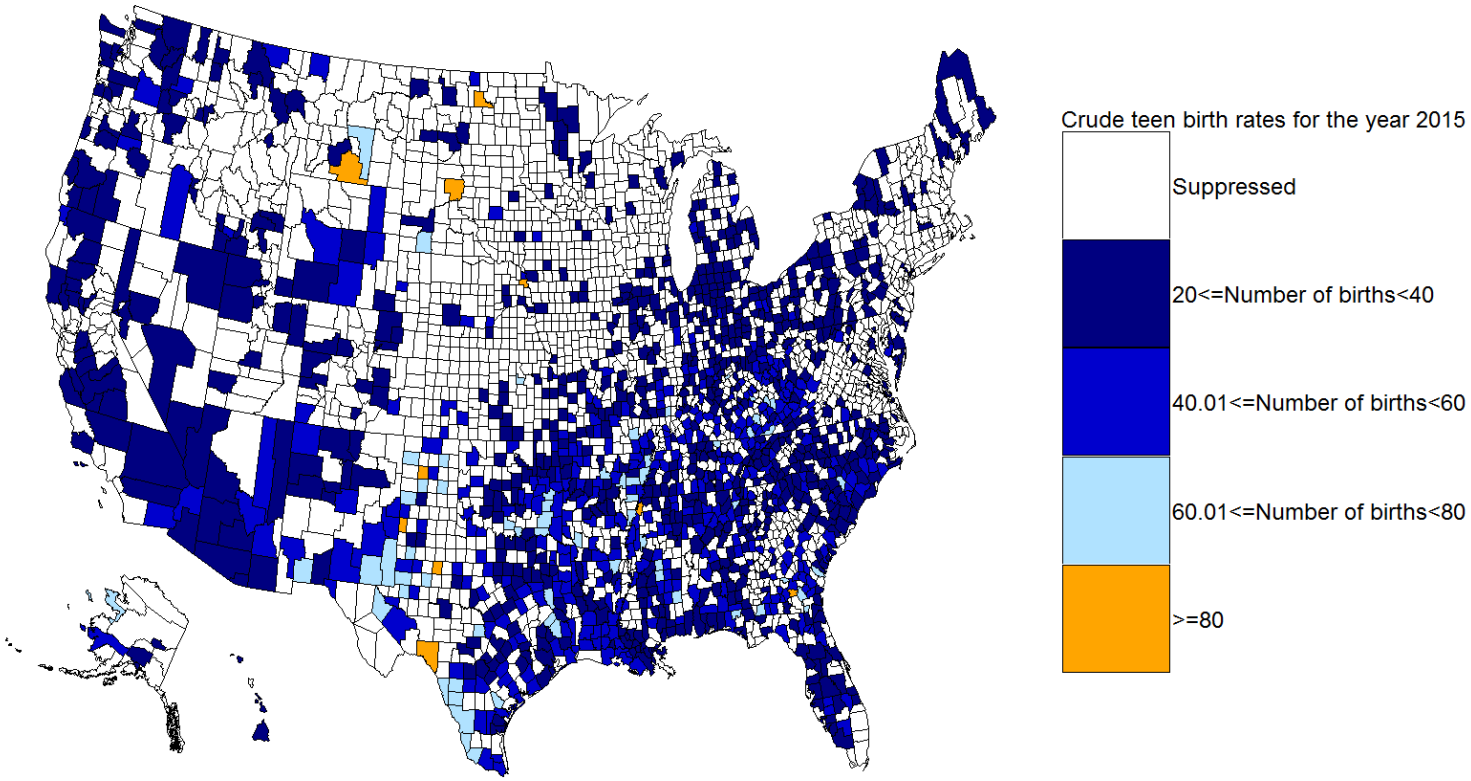


Figure: Teen birth rates based on less than 20 births for a county are suppressed for the year 2015.

Total number of counties by counts of teen births and percentage in the age group $15 - 19$ for years 2003-2015

	Equal to 0		Less than 10		Less than 20	
Year	Count	Percent	Count	Percent	Count	Percent
2003	56	1.78	512	16.31	929	29.60
2008	54	1.72	516	16.44	880	28.5
2015	102	3.25	797	25.4	1403	44.72

Table: Counts and percentage for total number of counties by counts of teen births in the age group $15 - 19$ reported to be equal to 0, less than 10, and less than 20 for 2003, 2008, and 2015 respectively.

Why Bayesian approach?

1. Main advantage: takes into account the uncertainty of the estimates/predictions
2. The inferential process accounts for spatial trend via spatially structured random effects: providing insight knowledge
3. Hierarchial model accounts for similarities based on neighbourhood structure
4. The space-time effects explain the differences/changes in time trend for different counties

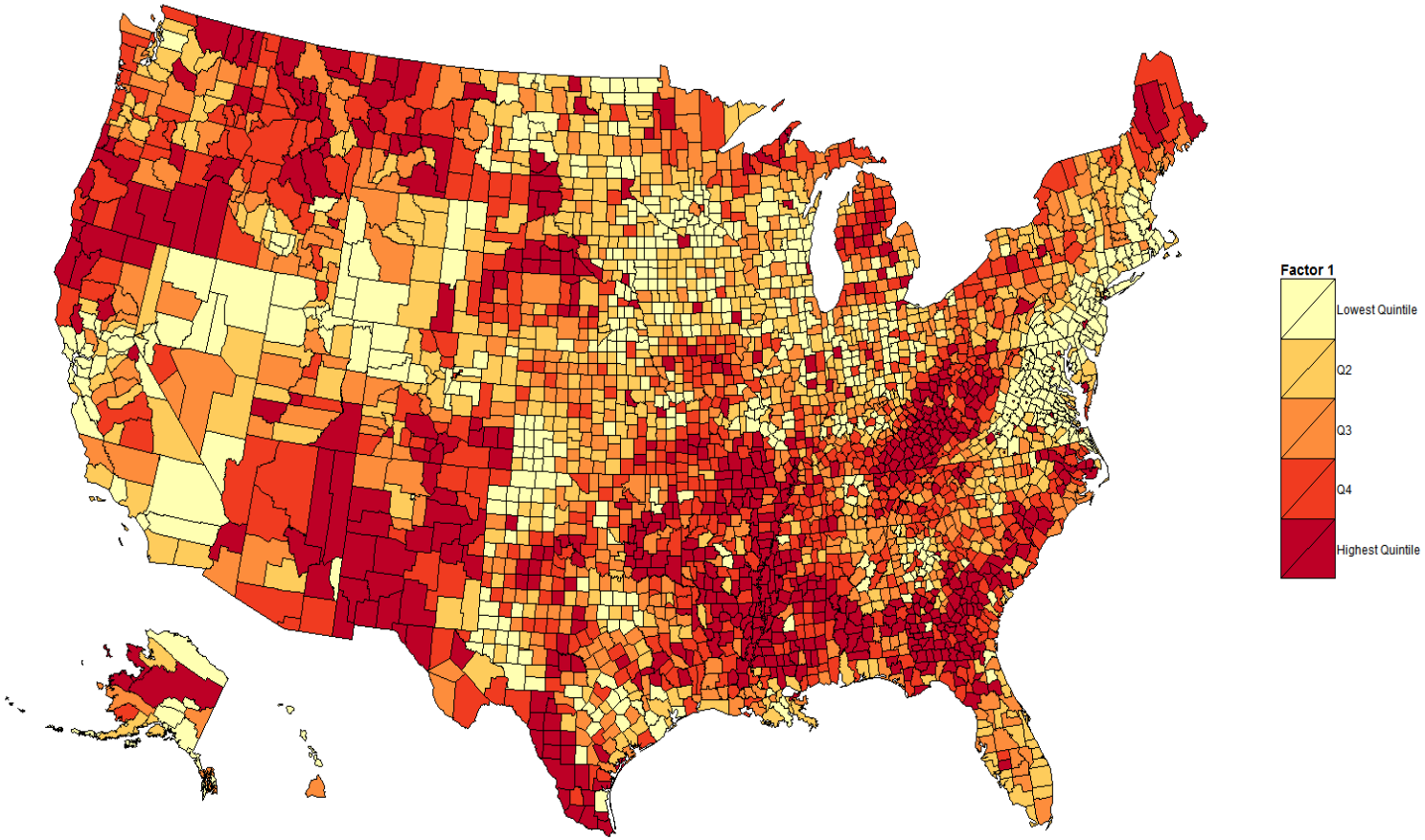
Bayesian model

- Define a spatial-temporal model using a hierarchical Bayesian framework
- Account for spatial and temporal trends
- Areas close to each other - more likely to share geographical characteristics related to the heath outcome
- Identification of temporal pattern: stronger for subsequent years than for years apart

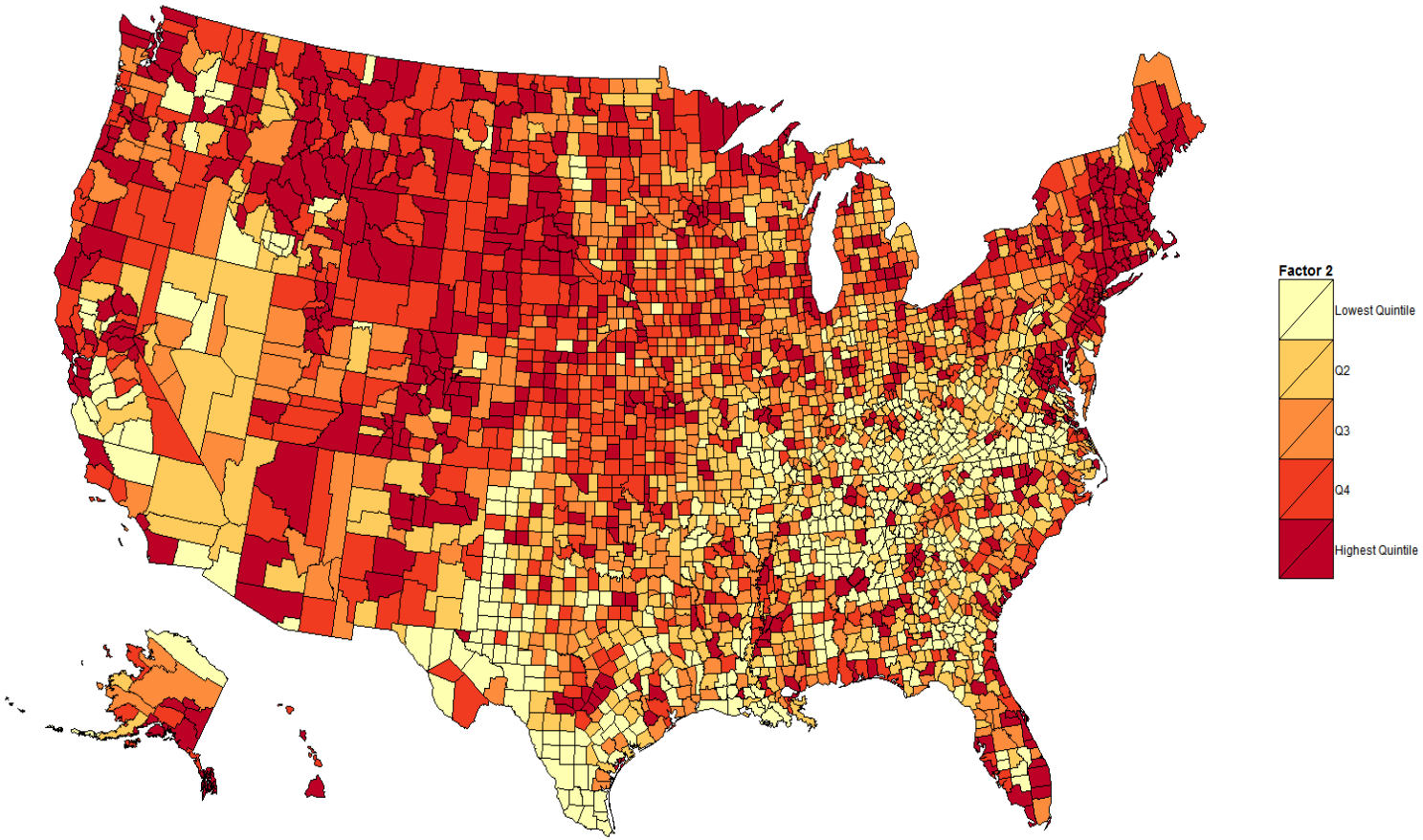
Covariates: Principal Component Analysis (PCA)

- Calculated the correlation between continuous covariates and the teen birth rate each year from 2003 - 2015
- Selected the covariates for which the absolute correlation was greater than 0.4 for a majority of the years
- There were 18 covariates selected
- Conducted PCA on these covariates
- Factors with an eigen value greater than 1 were retained, in this case 3 factors
- Varimax rotation was used

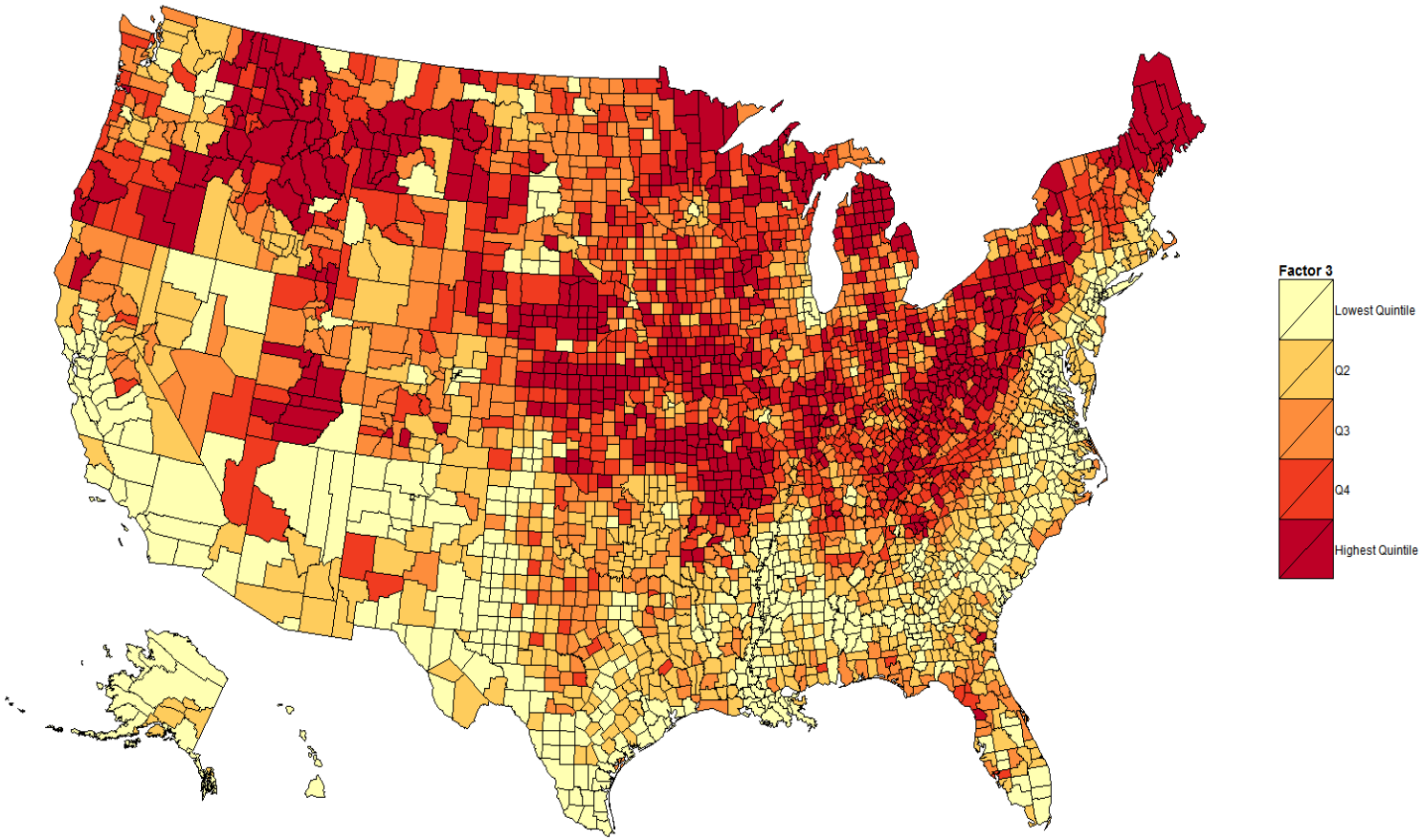
PCA - Construct 1 - High poverty and low income



PCA - Construct 2 - Educational level



PCA - Construct 3 - Race/Ethnicity: Percent White



Problem formulation

y_{it} : counts of births by county i and year t
n_{it} : counts of population by county i and year t
$y_{it} \sim Binomial(n_{it}, p_{it})$
p_{it} : probability of teen births in county i at time t
where $i = 1, \ldots, m$ areas and $t = 1, \ldots, T$ times.
General space-time model structure (Lawson, A. (2013)):
$logit(p_{it}) = \alpha_0 + A_i + B_t + C_{it}$,
where:
A_i : spatial group
B_t : temporal group
C_{it} : space-time interaction group

Proposed models

Model 1 With uncorrelated heterogeniety (non-spatial random effect)
$logit(p_{it}) = \alpha_0 + v_i$
Prior for $\alpha_0 \sim dflat$
Prior for v_i: $v_i \sim N(0, 1/\tau_v)$ termed as uncorrelated heterogeneity (variability)
τ_v is the precision
$Log(\tau_v) \sim LogGamma(1, 0.001)$

Proposed models

Model 2 Besag: convolution model
$logit(p_{it}) = \alpha_0 + u_i + v_i$
Intrinsic CAR prior for:$u_i|u_{-i}$
$u_i|u_{-i} \sim N(\bar{u}_i, r/n_{\delta_i})$ termed as correlated heterogeneity (variability), where
$u_{-i} = (u_1, u_2, \ldots, u_{i-1}, u_{i+1}, \ldots, u_m)$
n_{δ_i} : number of neighbors $= \sum_{j=1}^{m} \omega_{ij}$
δ_i : neighborhood of ith region
r : is the variance, $r = 1/\tau_r$:τ_r is the precision
$Log(\tau_r) \sim LogGamma(1, 0.001)$
$\bar{u}_i = \frac{\sum_{j=1}^{m} \omega_{ij} u_j}{\sum_{j=1}^{m} \omega_{ij}}$
$\alpha_0 \sim dflat$

Proposed models

Model 3 Random time effect (uncorrelated)
$logit(p_{it}) = \alpha_0 + u_i + v_i + \gamma 1_t$
$\gamma 2_t \sim N(0, 1/\tau_{\gamma 2})$ (uncorrelated)
$Log(\tau_{\gamma 2}) \sim LogGamma(1, 0.001)$

Proposed models

Model 4 Random time effect (uncorrelated) plus iid space-time interaction

$logit(p_{it}) = \alpha_0 + u_i + v_i + \gamma 2_t + \psi_{it}$

$\gamma 2_t \sim N(0, 1/\tau_{\gamma 2})$ (uncorrelated)

$Log(\tau_{\gamma 2}) \sim LogGamma(1, 0.001)$

$\psi_{i,t} \sim N(0, 1/\tau_{\psi})$ (uncorrelated)

$Log(\tau_{\psi}) \sim LogGamma(1, 0.001)$

Proposed models

Model 5 Space-time interaction (correlated)
$logit(p_{it}) = \alpha_0 + u_i + v_i + \gamma 2_t + \psi_{it}$
$\gamma 2_t \sim N(0, 1/\tau_{\gamma 2})$ (uncorrelated)
$Log(\tau_{\gamma 2}) \sim LogGamma(1, 0.001)$
$\psi_{i,t} \sim N(\psi_{i,t-1}, \tau_\psi)$ (randomwalk (Type 2 interaction))
$Log(\tau_\psi) \sim LogGamma(1, 0.001)$

Proposed models

Model 6 Random time effect (uncorrelated) plus space-time interaction (correlated) plus covariates
$logit(p_{it}) = \alpha_0 + u_i + v_i + \gamma 2_t + \psi_{it} + \mathbf{X}'\boldsymbol{\beta}$
$\gamma 2_t \sim N(0, 1/\tau_{\gamma 2})$ (uncorrelated)
$Log(\tau_{\gamma 2}) \sim LogGamma(1, 0.001)$
$\psi_{i,t} \sim N(\psi_{i,t-1}, \tau_\psi)$ (randomwalk (Type 2 interaction))
$Log(\tau_\psi) \sim LogGamma(1, 0.001)$
$\mathbf{X}$: covariates matrix X_i
$\boldsymbol{\beta}$: vector of regression parameters
$\beta \sim N(0, 100)$

Goodness of the fit: Deviance Information Criterion (DIC) and Watanabe-Akaike information criteria (WAIC)

Table: DIC, the effective number of parameters estimated and WAIC for the models via INLA.

Model	DIC	n.eff	WAIC
Model 1	472039	3083.18	489985
Model 2	471998.3	3026.625	489911.6
Model 3	285111.8	3037.31	287390.2
Model 4	271376	12347.61	271407.5
Model 5	267375.5	8406.752	268159.3
Model 6	**267251.3**	**8684.229**	**267528.2**

Model check: State direct TBRs vs INLA model based aggregated state level TBRs, 2003-2015

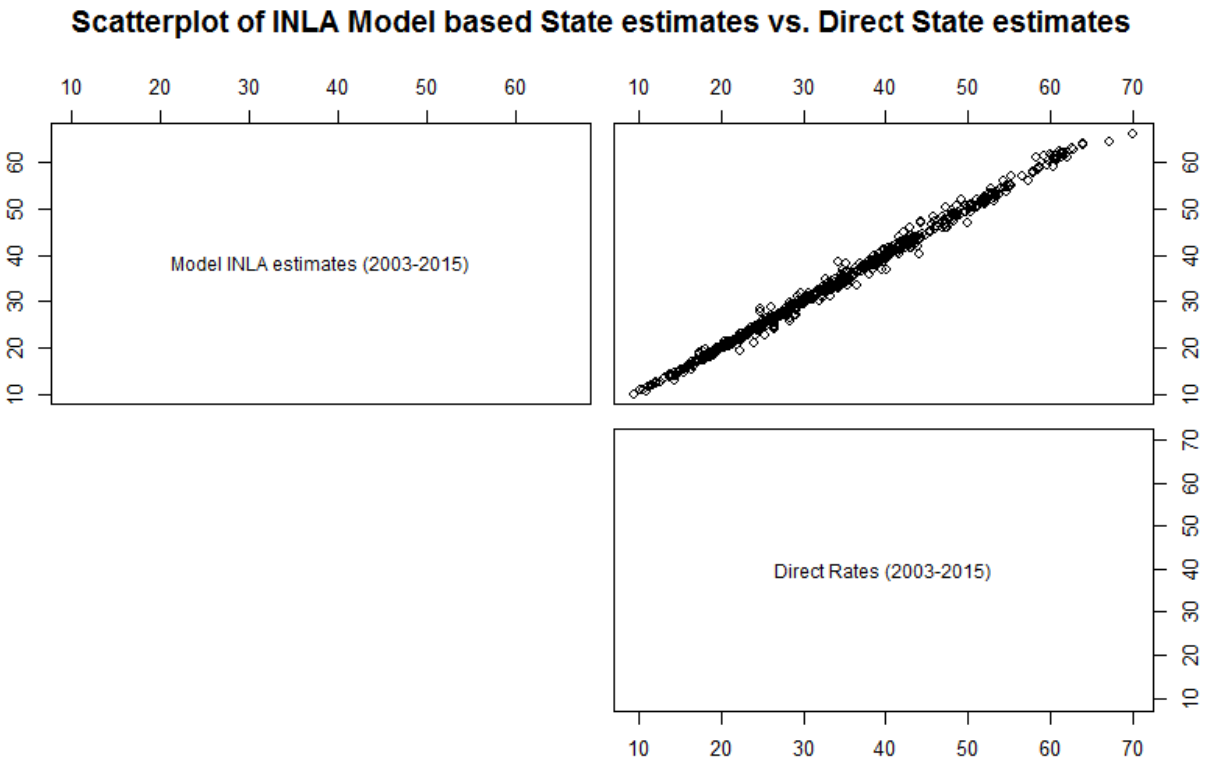


Figure: State direct TBRs vs INLA based state TBRs 2003-2015.

Model accuracy: MCMC vs INLA TBRs. **Computation time: INLA (24 hours) vs MCMC (9 weeks)**

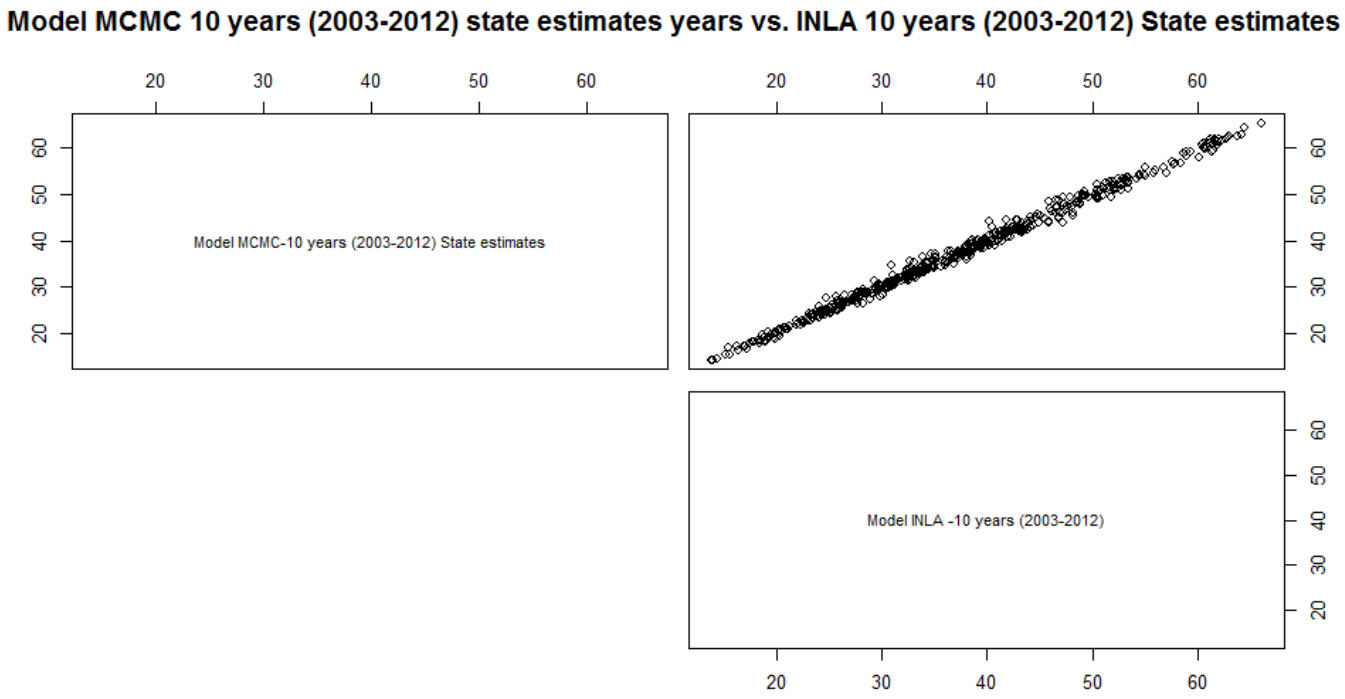
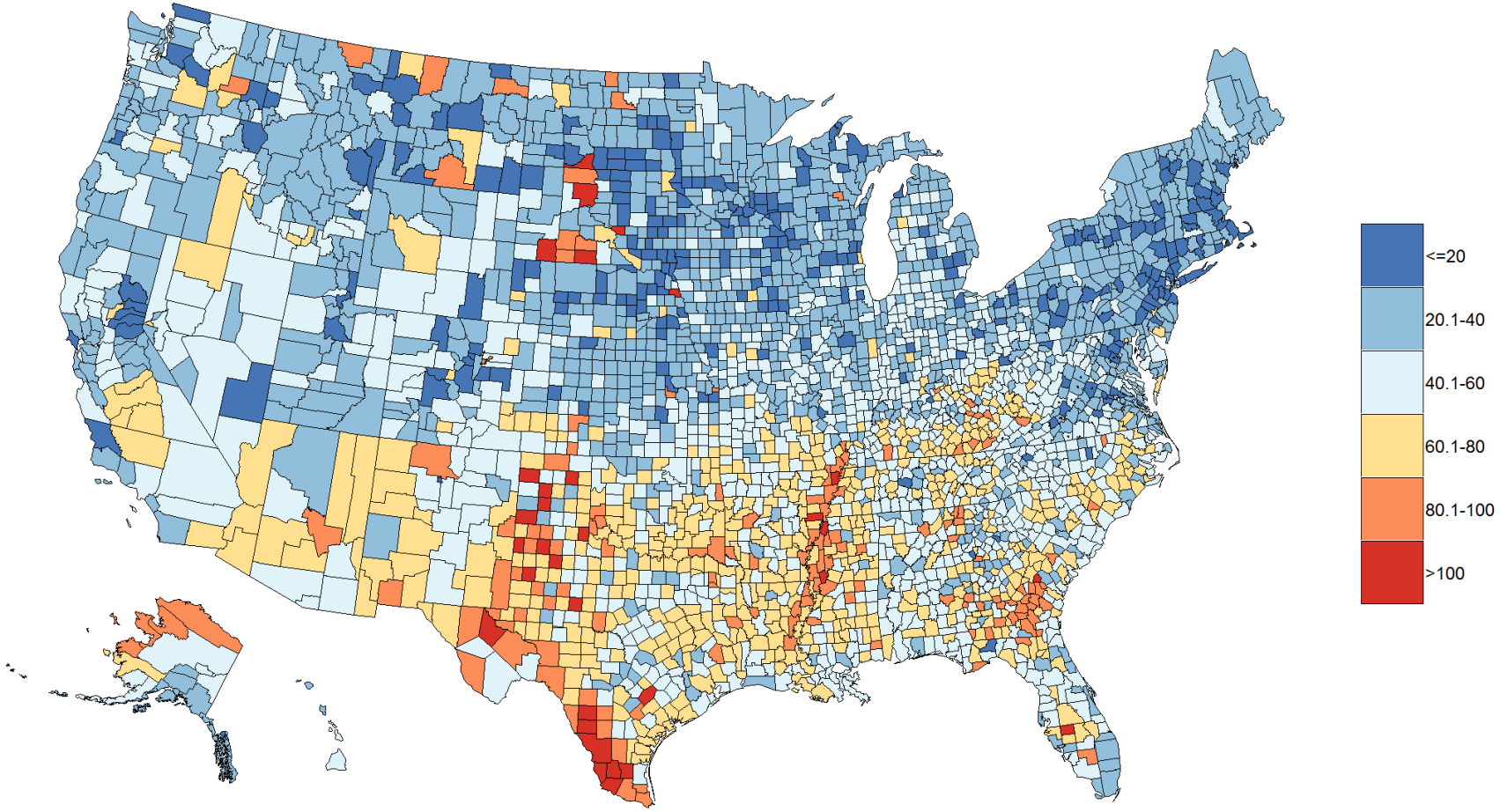
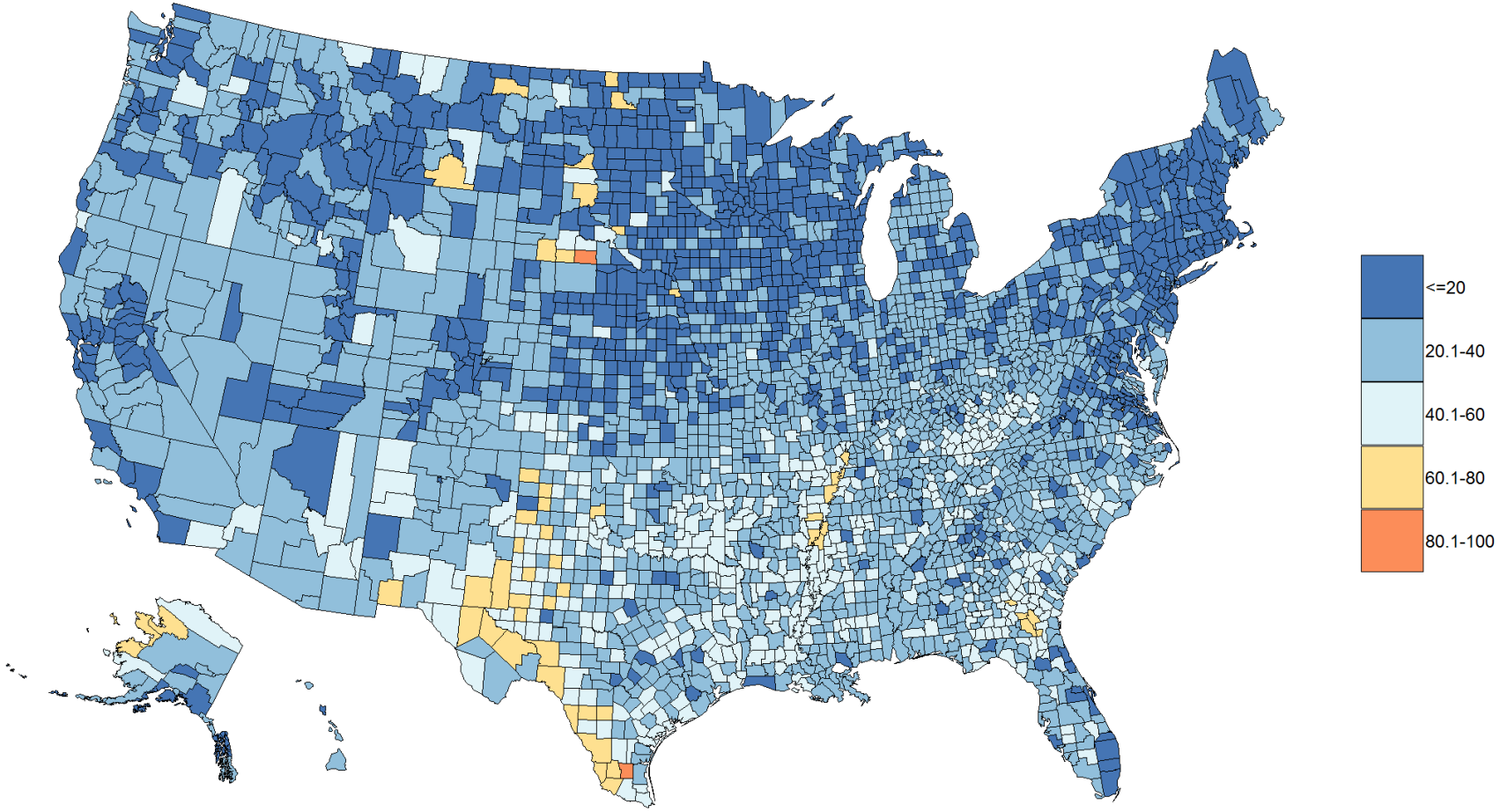


Figure: MCMC based vs INLA based aggregated state level TBRs, 2003-2012.

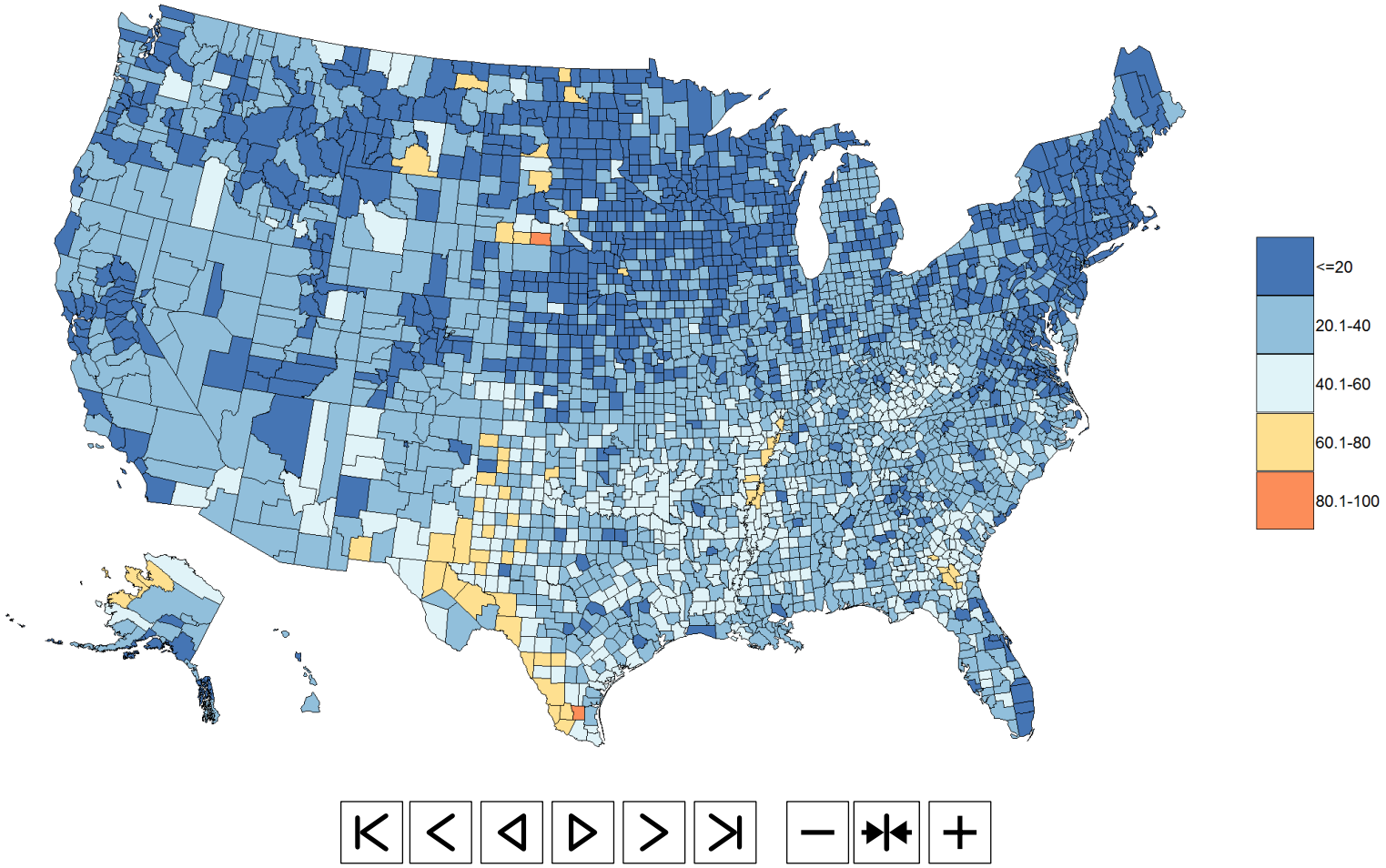
Predicted TBRs 2003



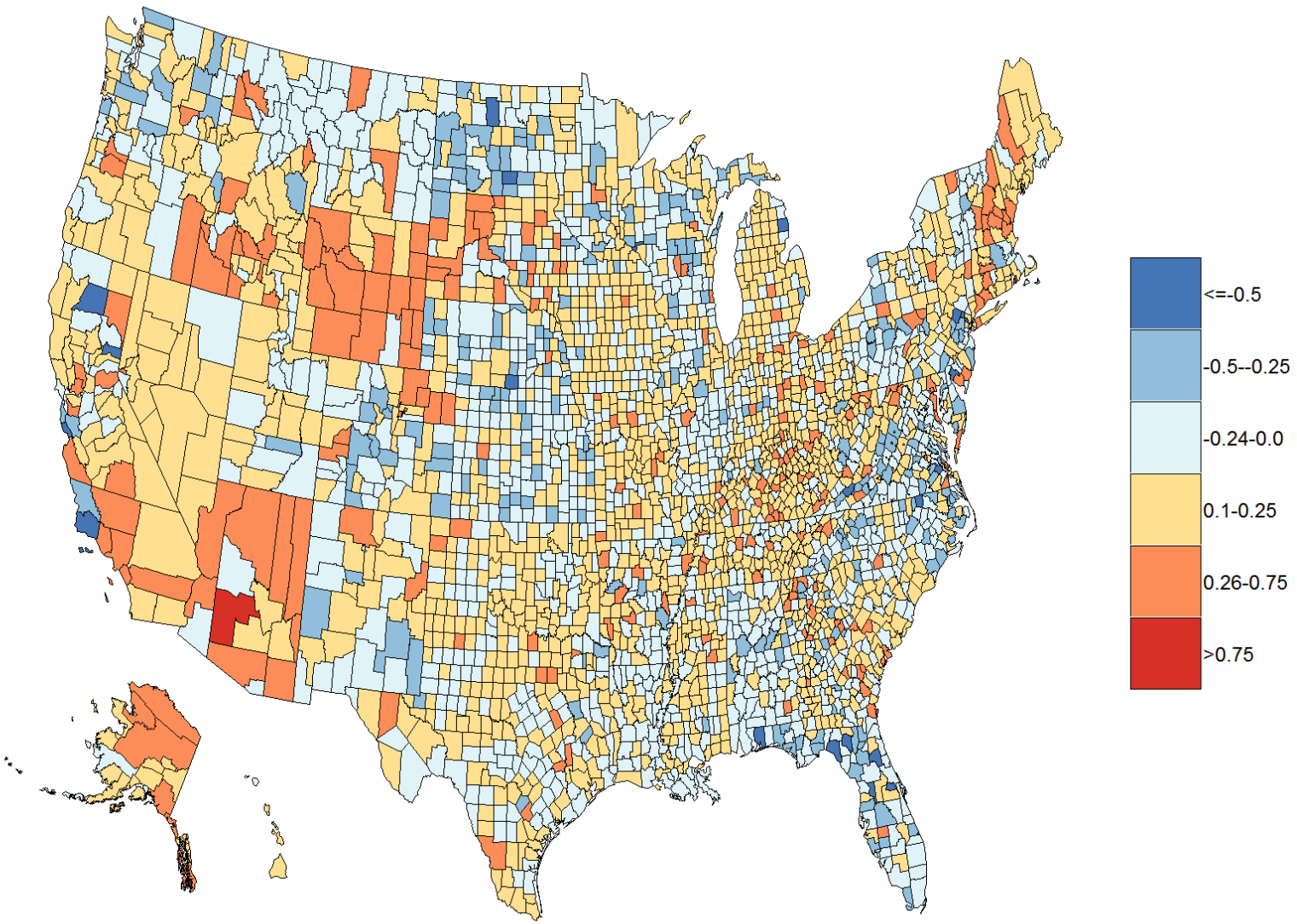
Predicted TBRs 2015



Animation of the predicted TBRs: 2003 through 2015



Animation of Type II interaction effects: 2003-2015



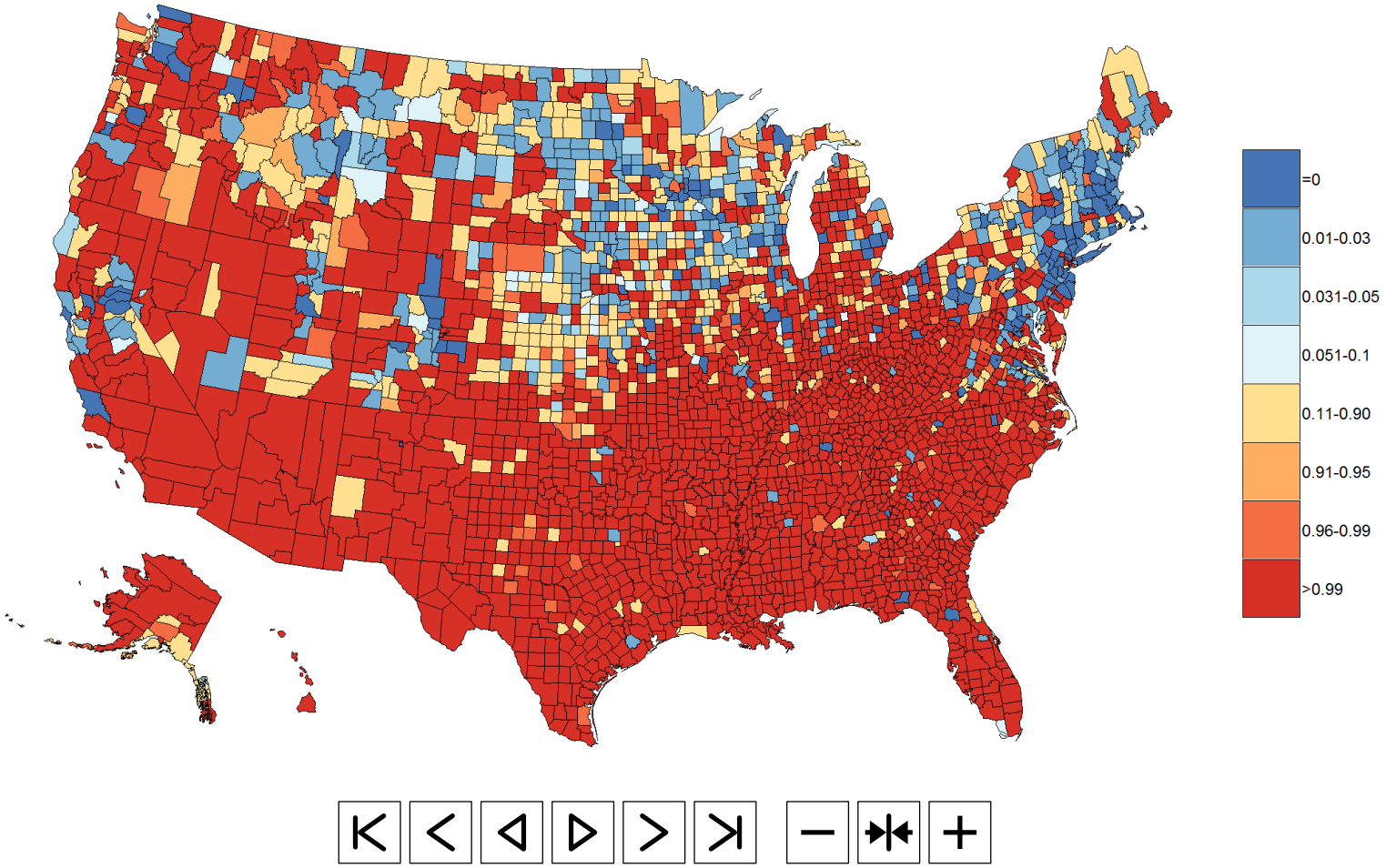
Exceedance probabilities

1. The probability that the predicted TBRs exceed a certain threshold
2. **Can be used to quantify the uncertainty associated with the county level TBRs. Example: Probability that the predictions exceed the mean county level TBRs**

Exceedance probabilities: threshold

Probability of exceeding a threshold
Threshold set as the mean crude teen birth rate for the year 2015

Animation of Exceedance probabilities: 2003 through 2015



Trends in TBRs: uncorrelated time effect $\gamma 2_t$ and the 95 percent Bayesian credible intervals

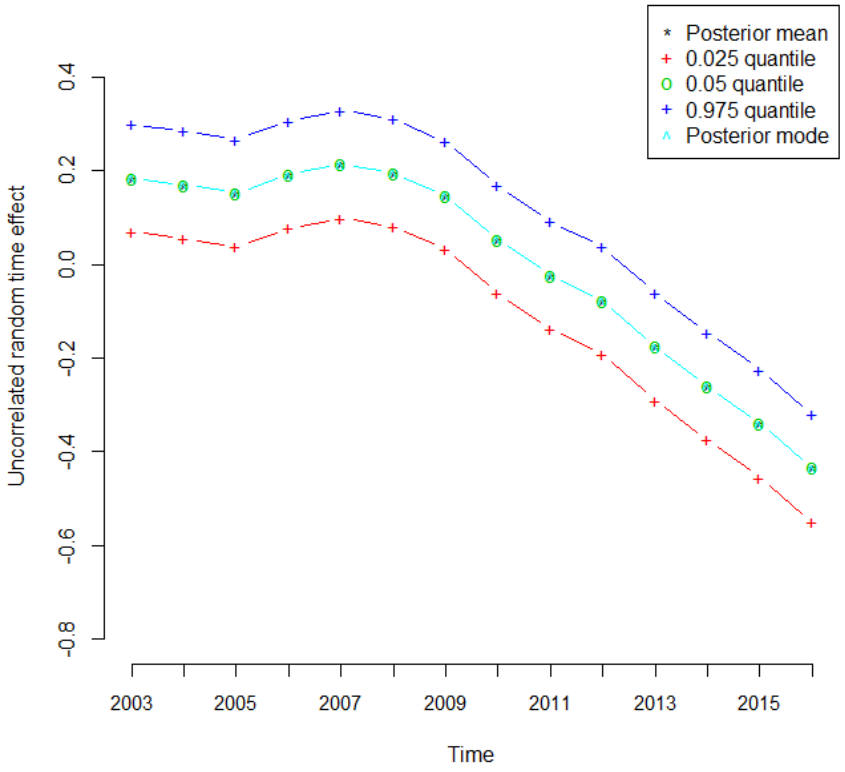


Figure: Uncorrelated time effect $\gamma 2_t$.

Correlated heterogeneity

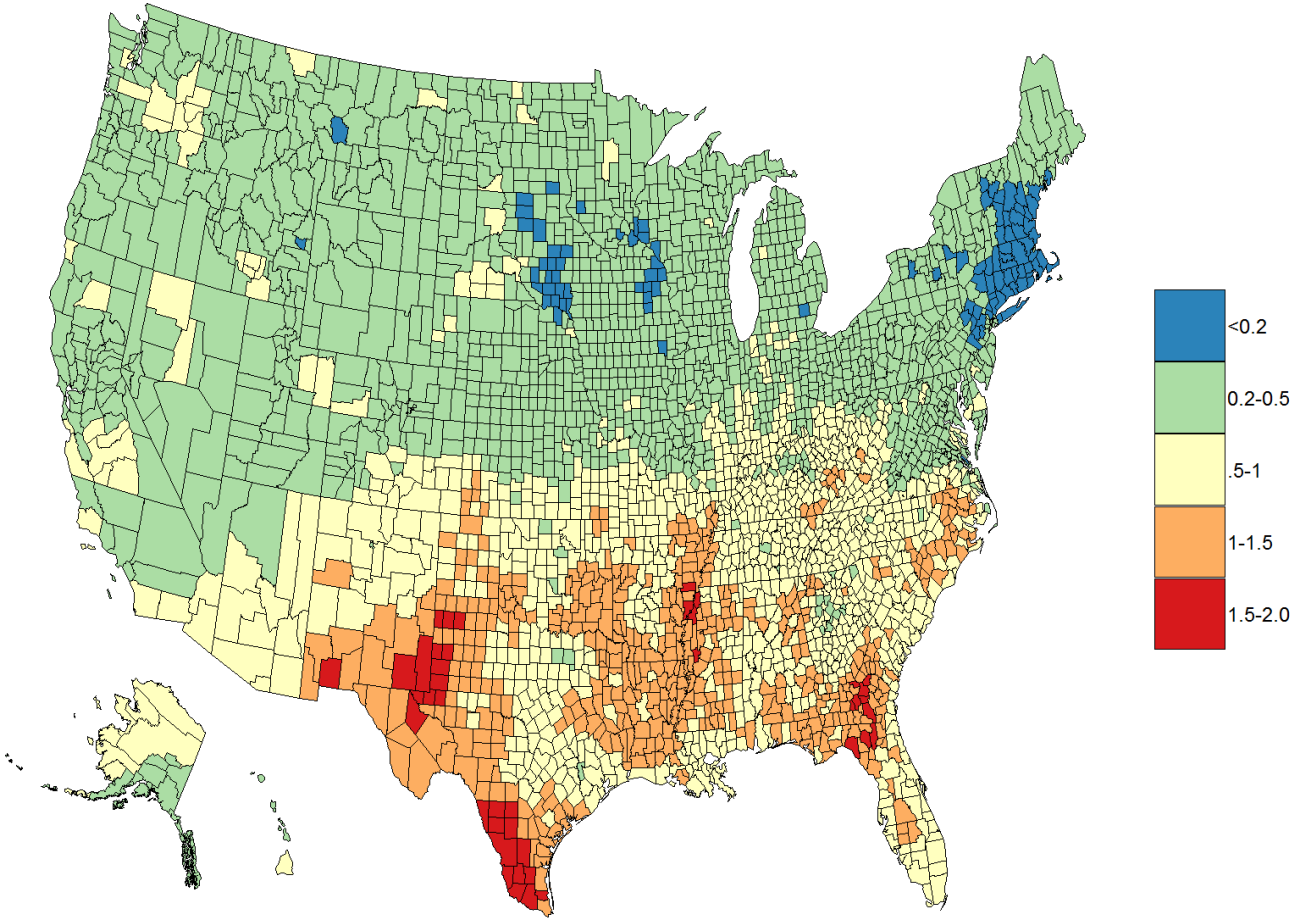
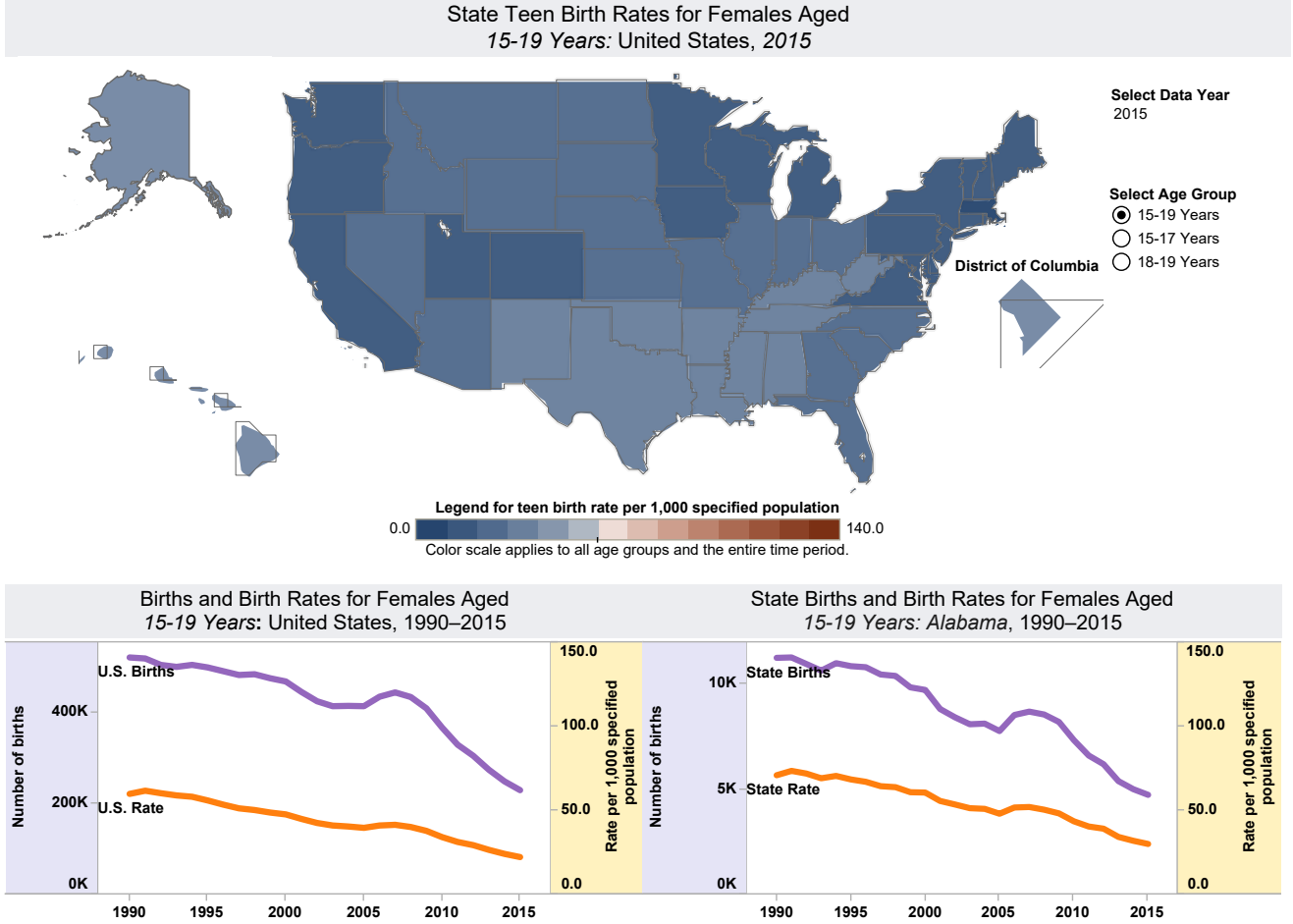


Figure: Correlated heterogeneity u_i.

Data visualization: Trends in teen birth rates

1. Data visualization done in the software Tableau
2. State level: Interactive maps and trend lines for teen birth rates (from birth certificates) for females aged 15-19, 15-17, and 17-19 for each of the 50 states
3. Can zoom in on individual states

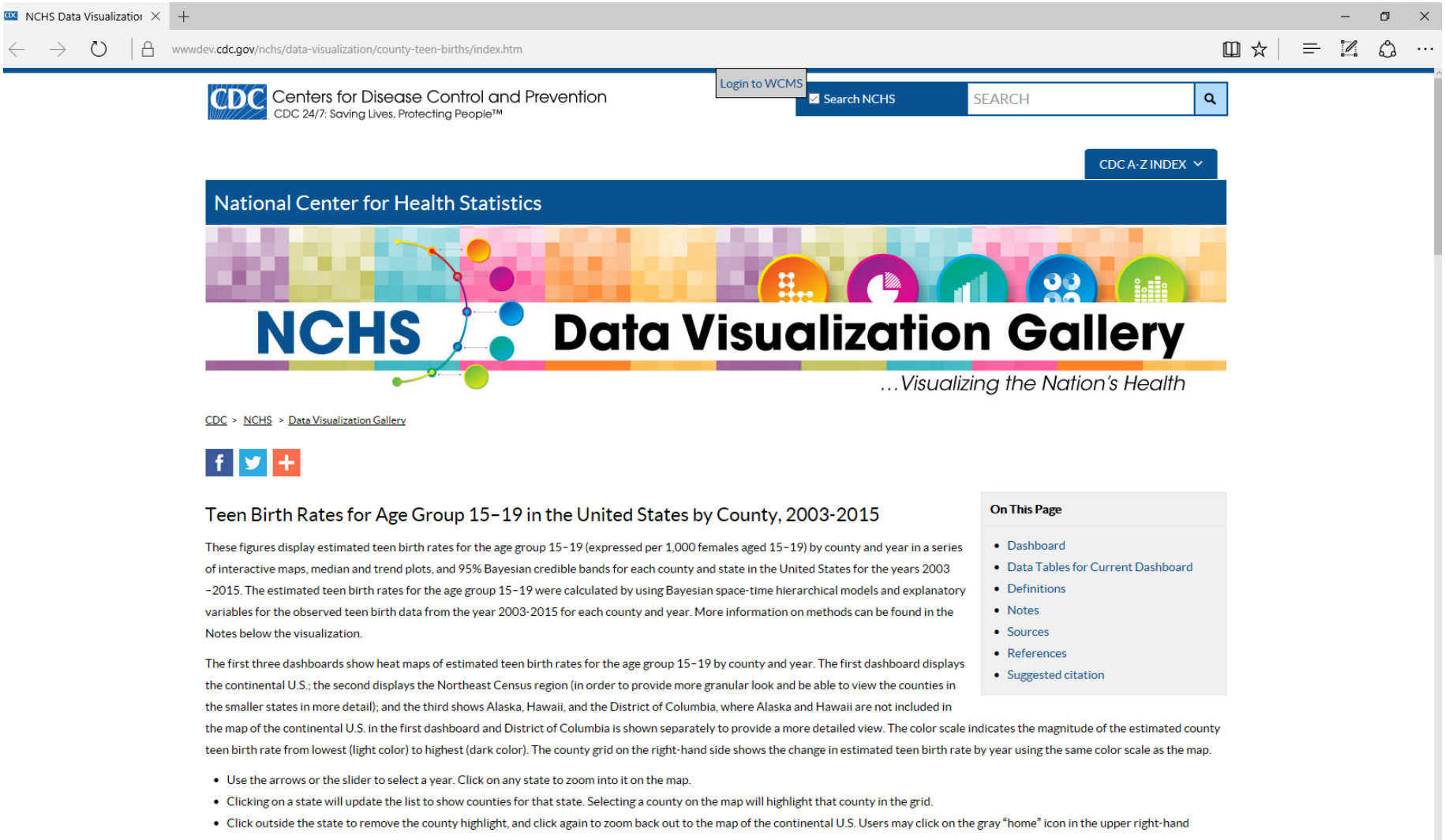
Data visualization: State level teen birth rates: 1990-2015



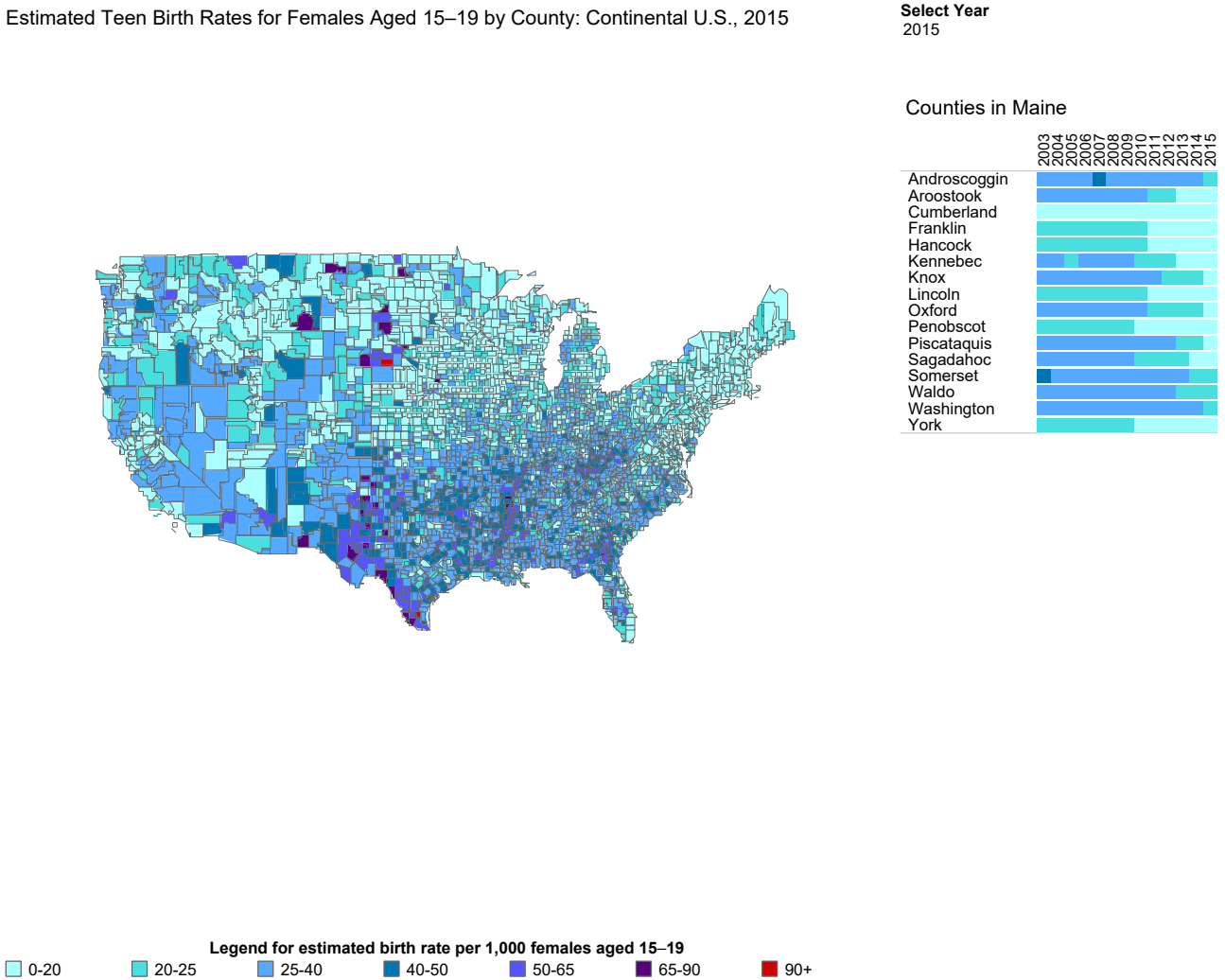
Data visualization: County level teen birth rates, 2003-2015

1. Contains four dashboards
2. County level: Interactive maps and trend lines for estimated teen birth rates for females aged 15-19 for 3137 counties
3. Trend lines, geographic variation and 95 percent Bayesian credible bands for the years 2003-2015 (soon to be updated with 2016)
4. Can zoom in on individual counties for more granular look
5. https://www.cdc.gov/nchs/data-visualization/county-teen-births/

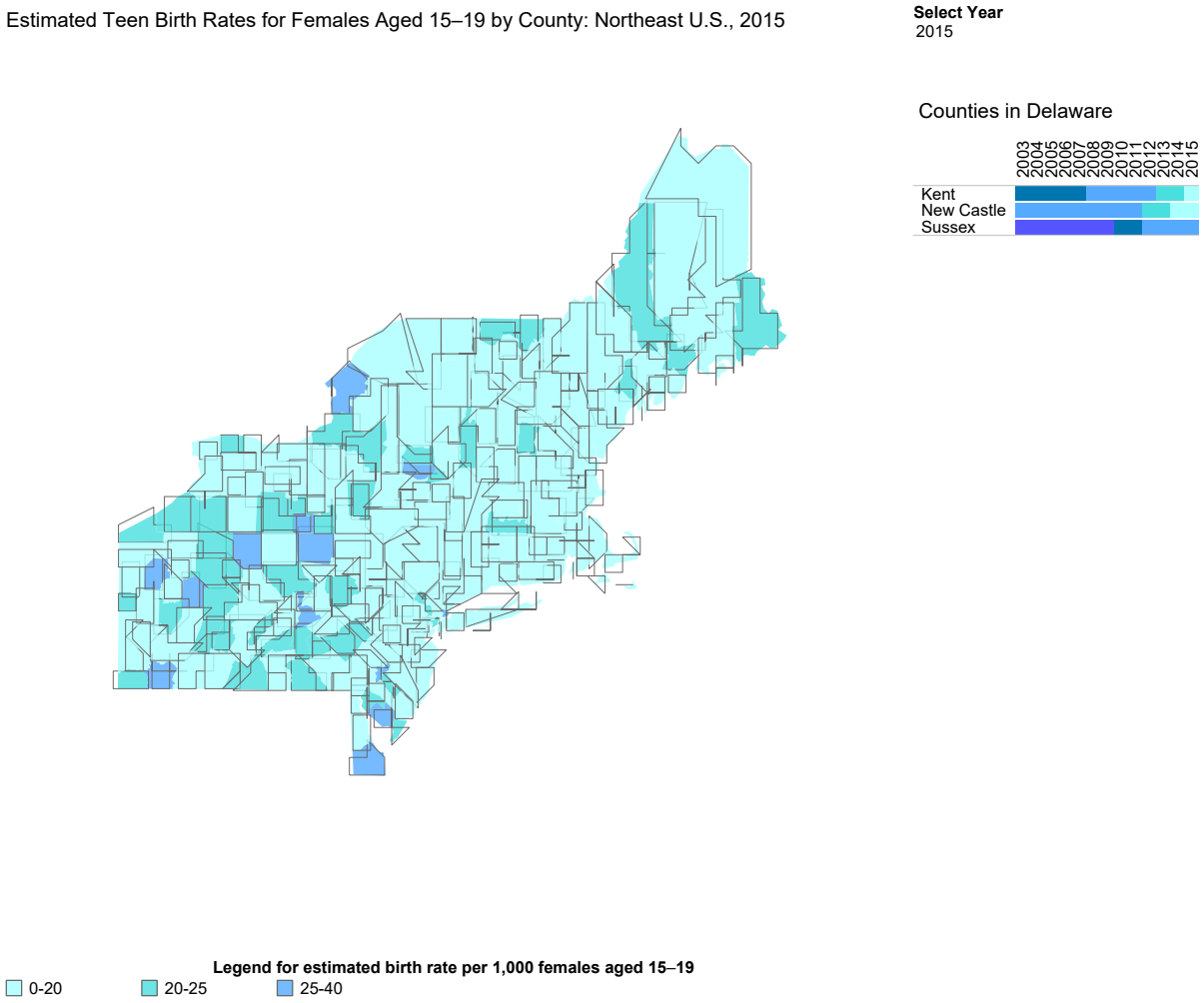
Data visualization gallery



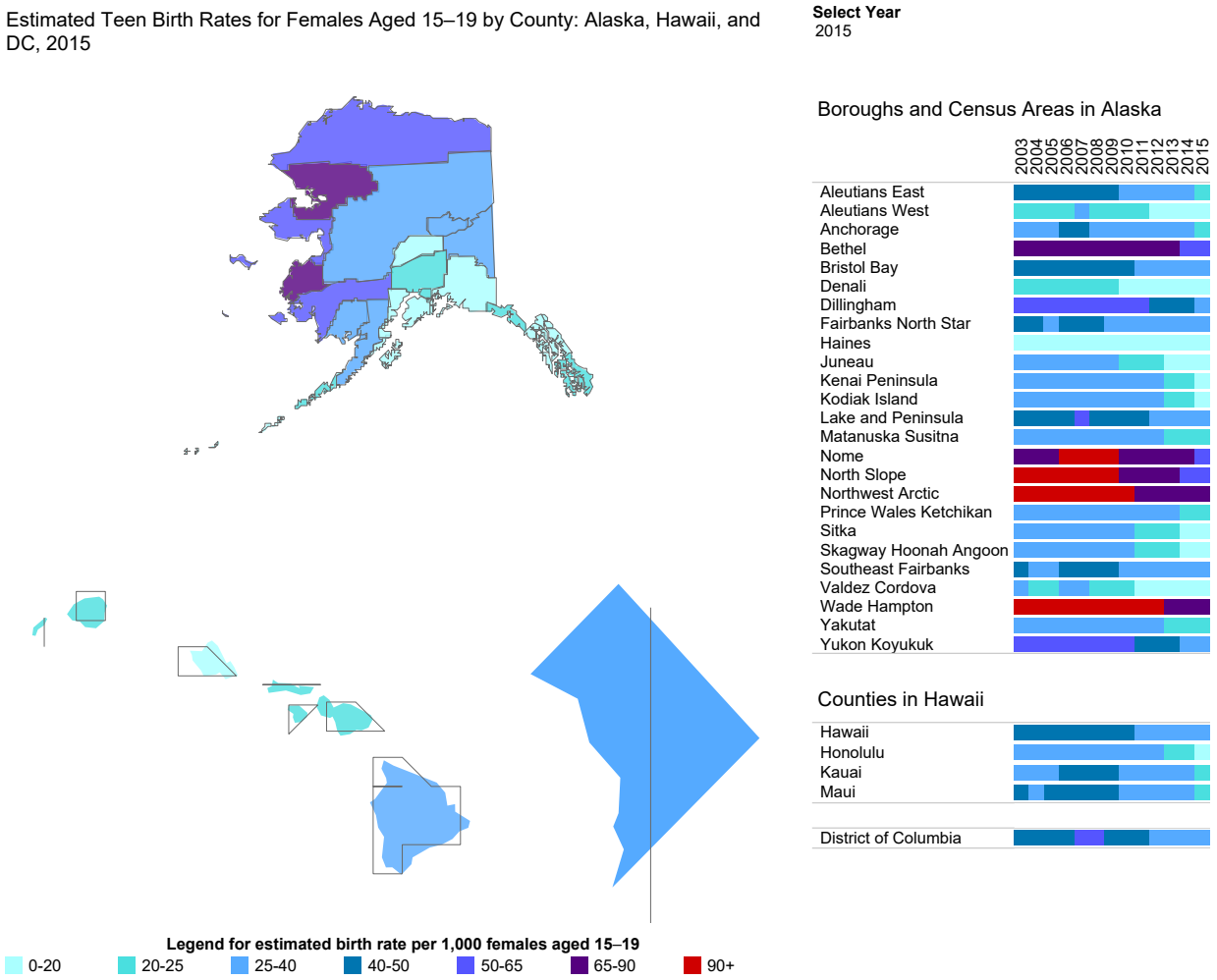
Data visualization



Data visualization

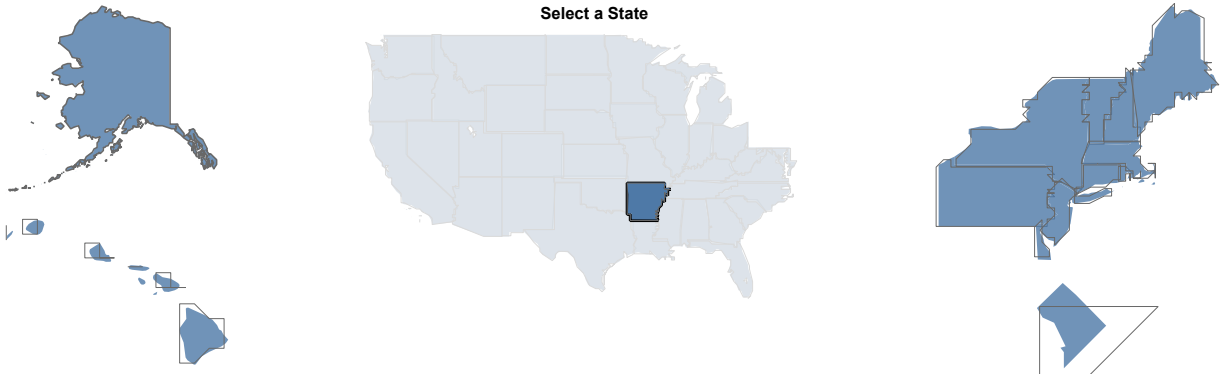


Data visualization

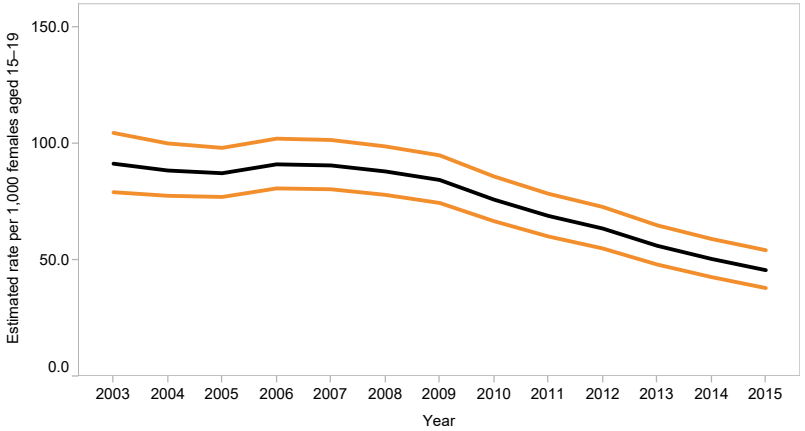


Data visualization

County Estimated Teen Birth Rate Medians and Credible Intervals



Estimated Teen Birth Rates for Females Aged 15–19 for Counties in Arkansas



Select a County from the table below to display it in the Credible Intervals line graph.

	Estimated Teen Birth Rate	Lower Confidence Limit	Upper Confidence Limit
Calhoun	46.4	36.0	58.4
Carroll	72.2	62.8	82.4
Chicot	73.6	62.4	85.9
Clark	33.9	28.9	39.4
Clay	63.2	53.7	73.7
Cleburne	52.8	45.1	61.3
Cleveland	56.6	46.0	68.4
Columbia	48.9	42.3	56.2
Conway	65.8	56.7	75.6
Craighead	53.0	47.8	58.4
Crawford	63.0	56.4	70.1
Crittenden	90.0	81.6	98.8
Cross	68.6	59.1	78.9
Dallas	58.2	47.8	69.8
Desha	91.5	79.3	104.7
Drew	49.5	42.2	57.4
Faulkner	35.5	31.9	39.3
Franklin	56.7	48.3	66.0

Select a View

- ○ Median of County Rates
- ◉ Credible Intervals

Legend for measures for estimated teen birth rate per 1,000 females

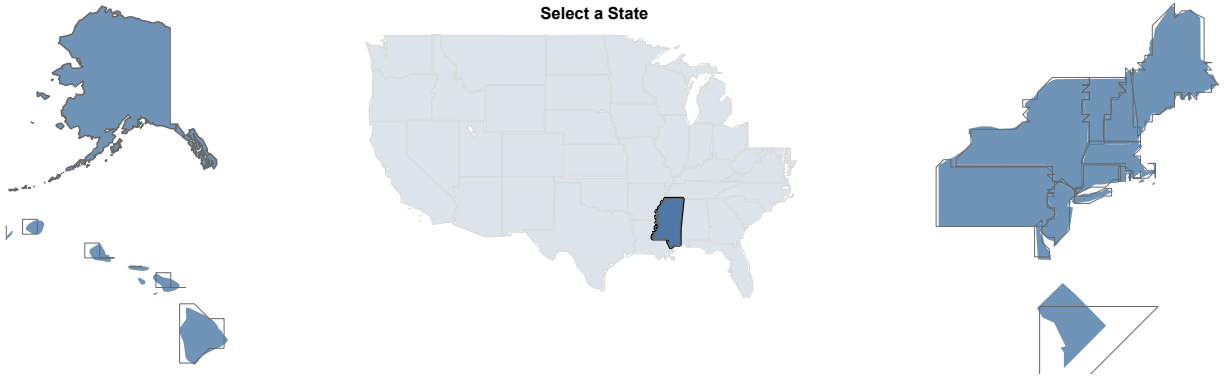
- Upper Confidence Limit
- Lower Confidence Limit
- Estimated Teen Birth Rate

Select a year to display estimated teen birth rates and 95% Bayesian credible bands for females aged 15–19 in the above table.

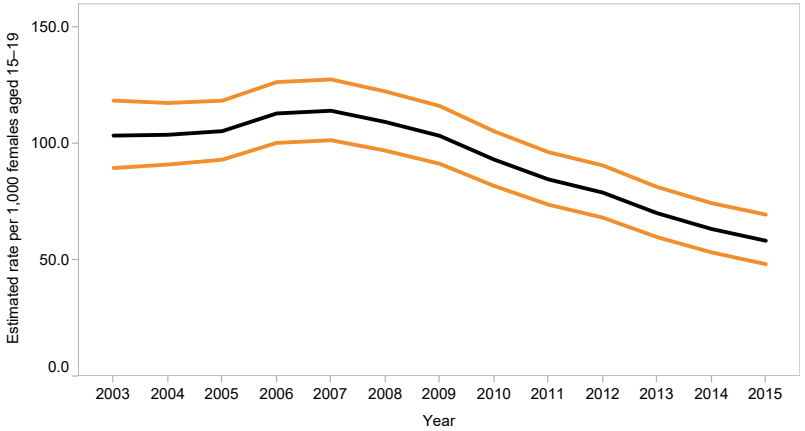
2003

Data visualization

County Estimated Teen Birth Rate Medians and Credible Intervals



Estimated Teen Birth Rates for Females Aged 15–19 for Counties in Mississippi



Select a County from the table below to display it in the Credible Intervals line graph.

	Estimated Teen Birth Rate	Lower Confidence Limit	Upper Confidence Limit
Clarke	58.8	50.1	68.4
Clay	60.4	52.2	69.2
Coahoma	91.2	81.7	101.3
Copiah	60.9	53.7	68.7
Covington	73.6	64.0	83.9
De Soto	46.6	42.2	51.1
Forrest	48.9	44.1	53.9
Franklin	50.5	41.1	61.2
George	84.1	73.3	95.8
Greene	61.0	50.9	72.2
Grenada	65.1	56.5	74.5
Hancock	53.3	47.0	60.2
Harrison	64.1	59.5	68.8
Hinds	60.1	56.3	64.0
Holmes	79.3	69.9	89.4
Humphreys	103.6	89.6	118.6
Issaquena	74.2	53.9	97.8
Itawamba	48.9	41.9	56.4

Select a year to display estimated teen birth rates and 95% Bayesian credible bands for females aged 15–19 in the above table.

2003

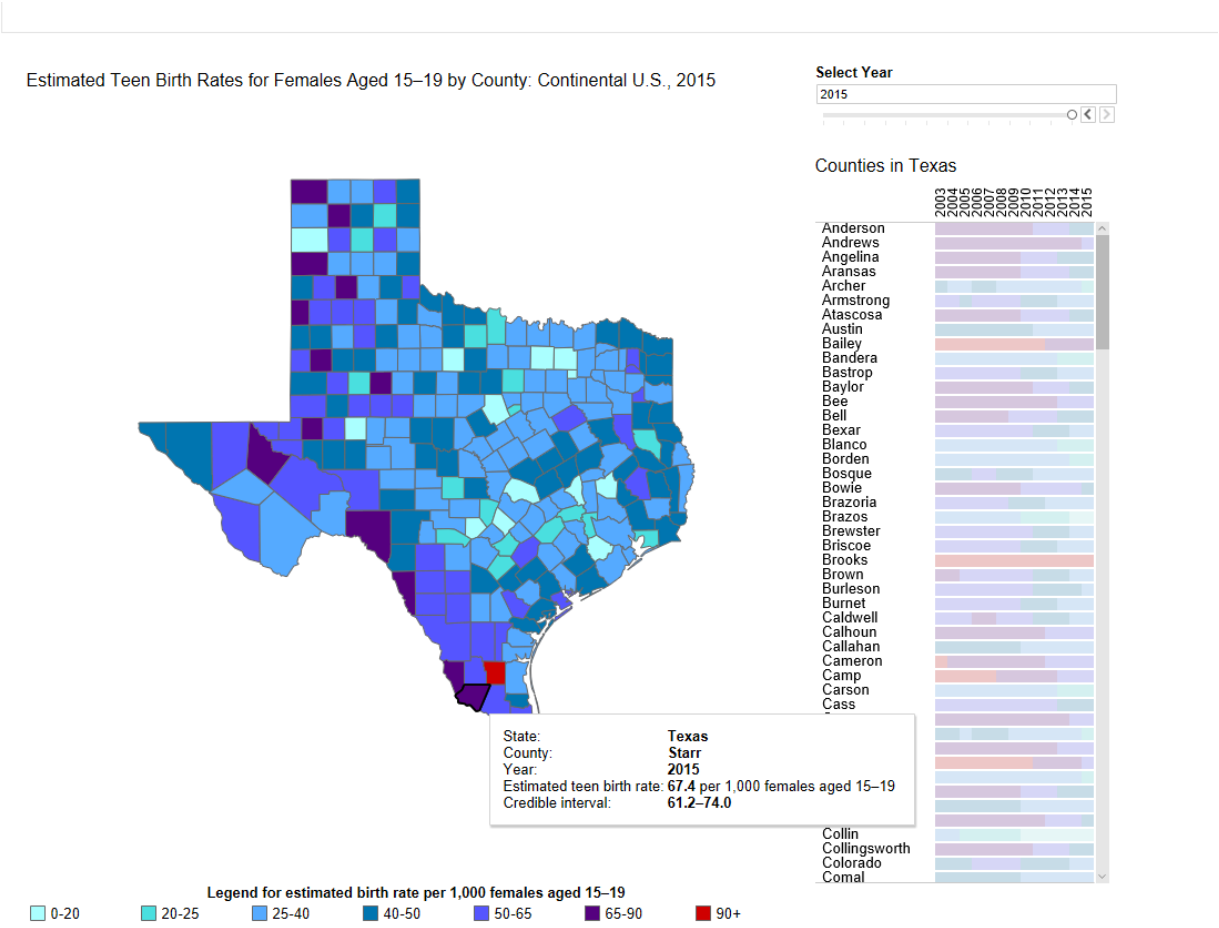
Select a View

- ○ Median of County Rates
- ◉ Credible Intervals

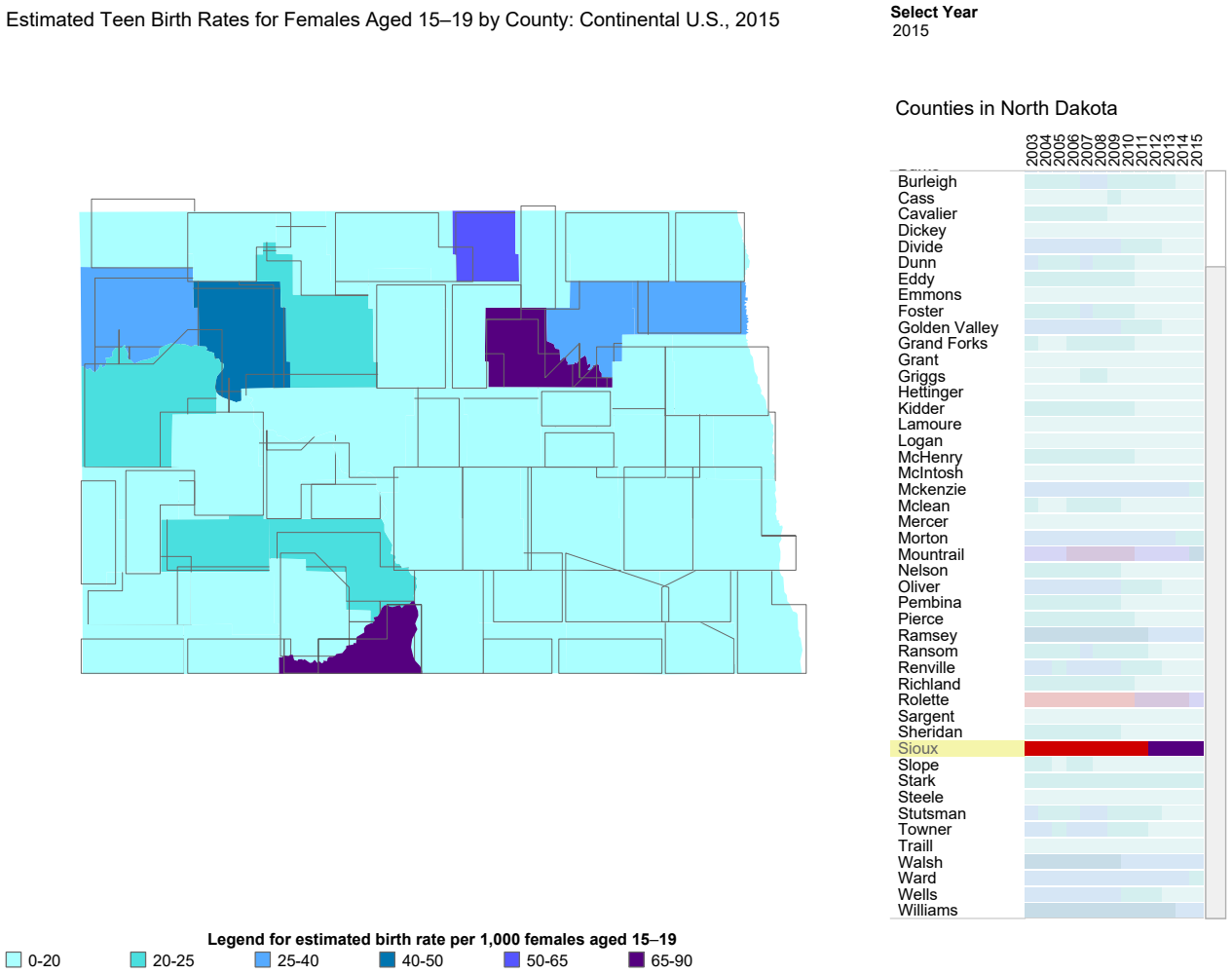
Legend for measures for estimated teen birth rate per 1,000 females

- Upper Confidence Limit
- Lower Confidence Limit
- Estimated Teen Birth Rate

Data visualization



Data visualization



Data visualization

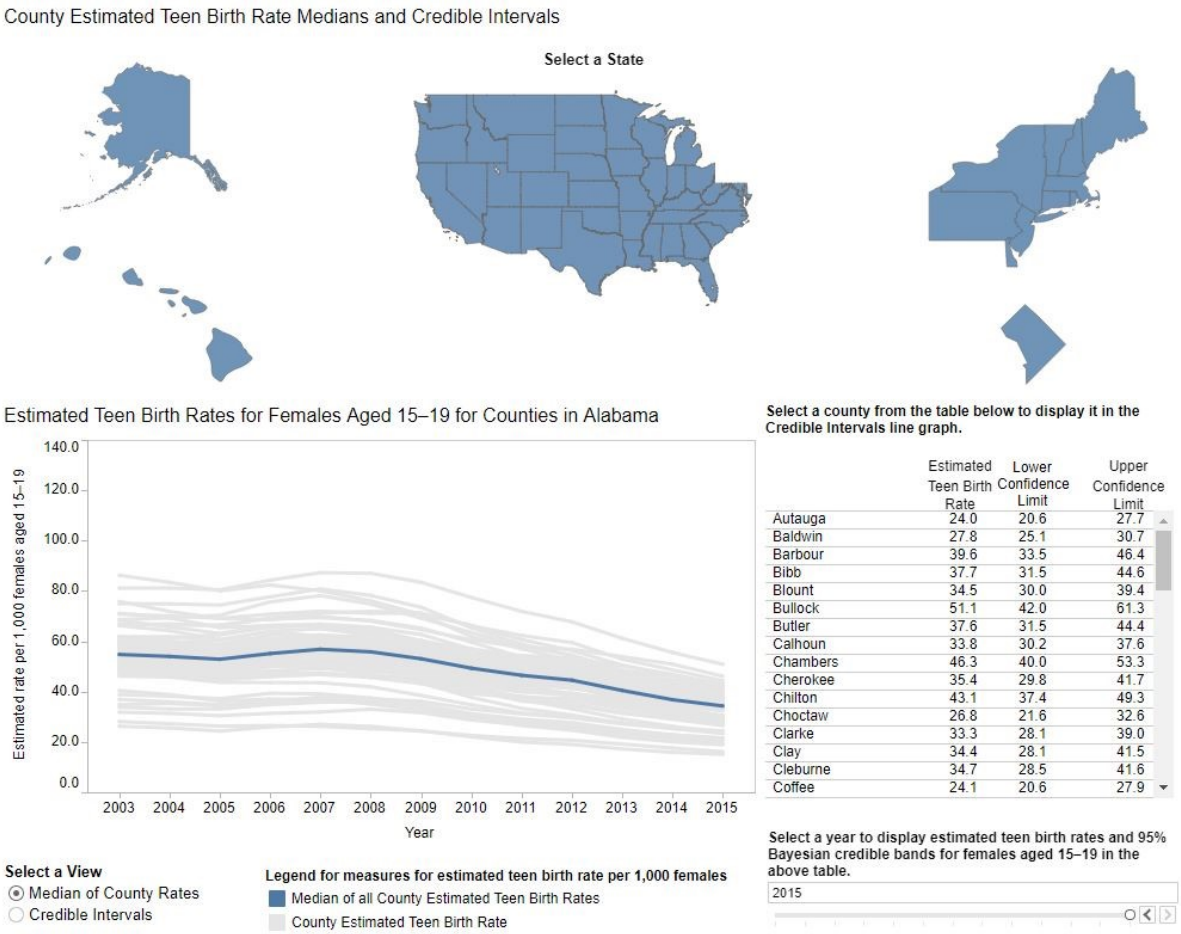


Figure: Estimated median teen birth rates (per 1,000) for counties in Alabama.

Small area direct estimates: example of county level suicide rates (SRs)

1. Age adjusted suicide rates increased by 27 percent from 1999-2015 (10.5-13.3 per 100,000 population)
2. County level estimates for less than 20 events are suppressed
3. Unreliable: for example rural areas fewer suicides and small population sizes
4. Past studies aggregate over several years (for example WISQARS (2008-2014)) may mask temporal trends
5. Aggregation of counties to produce larger geographic areas: may mask urban-rural differences and sub-state variations

Total number of counties by counts of suicides for years 2005-2015

	Equal to 0		Less than 10		Less than 20	
Year	Count	Percent	Count	Percent	Count	Percent
2005	475	15.12	2405	76.59	2775	88.37
2009	427	13.6	2349	74.8	2716	86.5
2015	360	11.5	2186	69.6	2646	84.3

Table: Counts and percentage for total number of counties by counts of suicides reported to be equal to 0, less than 10, and less than 20 for each 2005, 2009, and 2015 respectively.

WISQARS: aggregation in time (2008-2014)

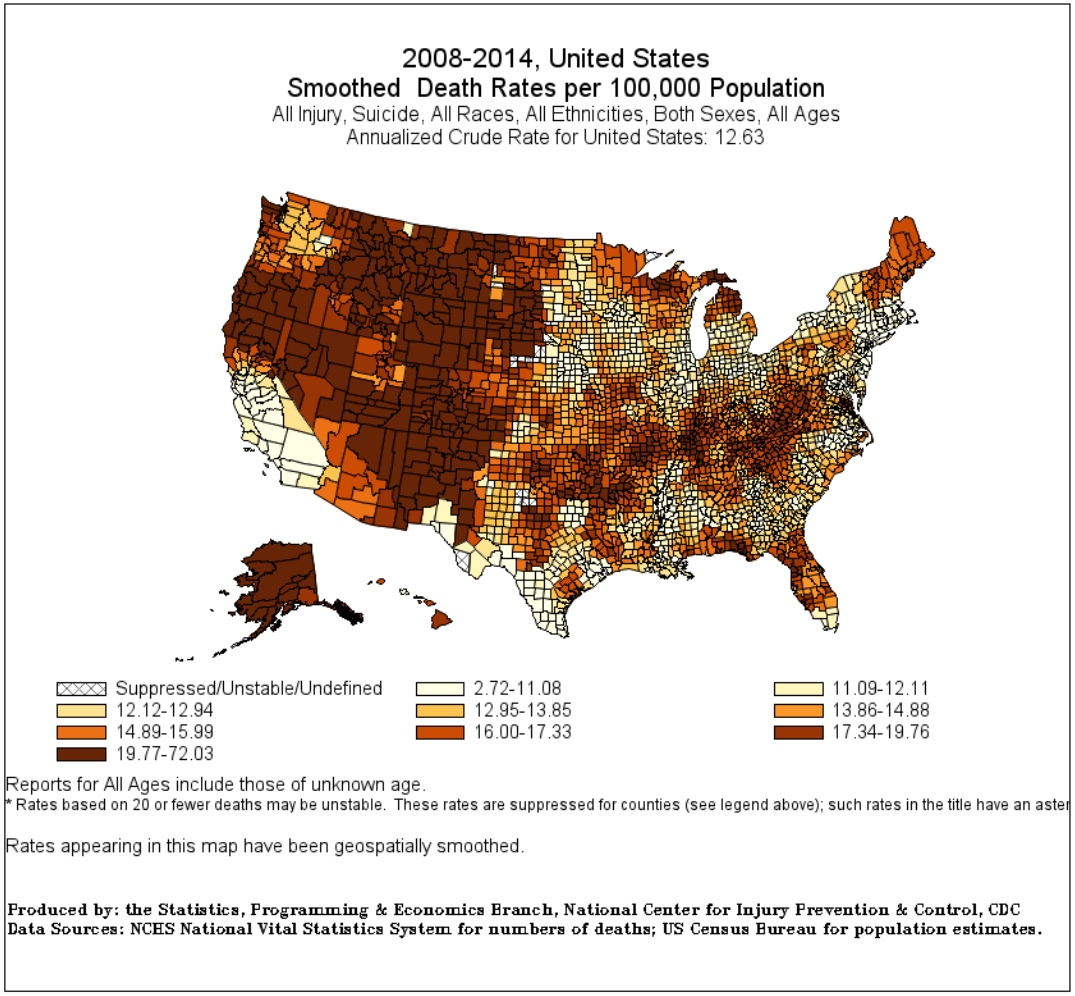


Figure: Geographic variation for aggregated suicide rates over 7 years.

Data suppression: Crude suicide number of deaths for the year 2015

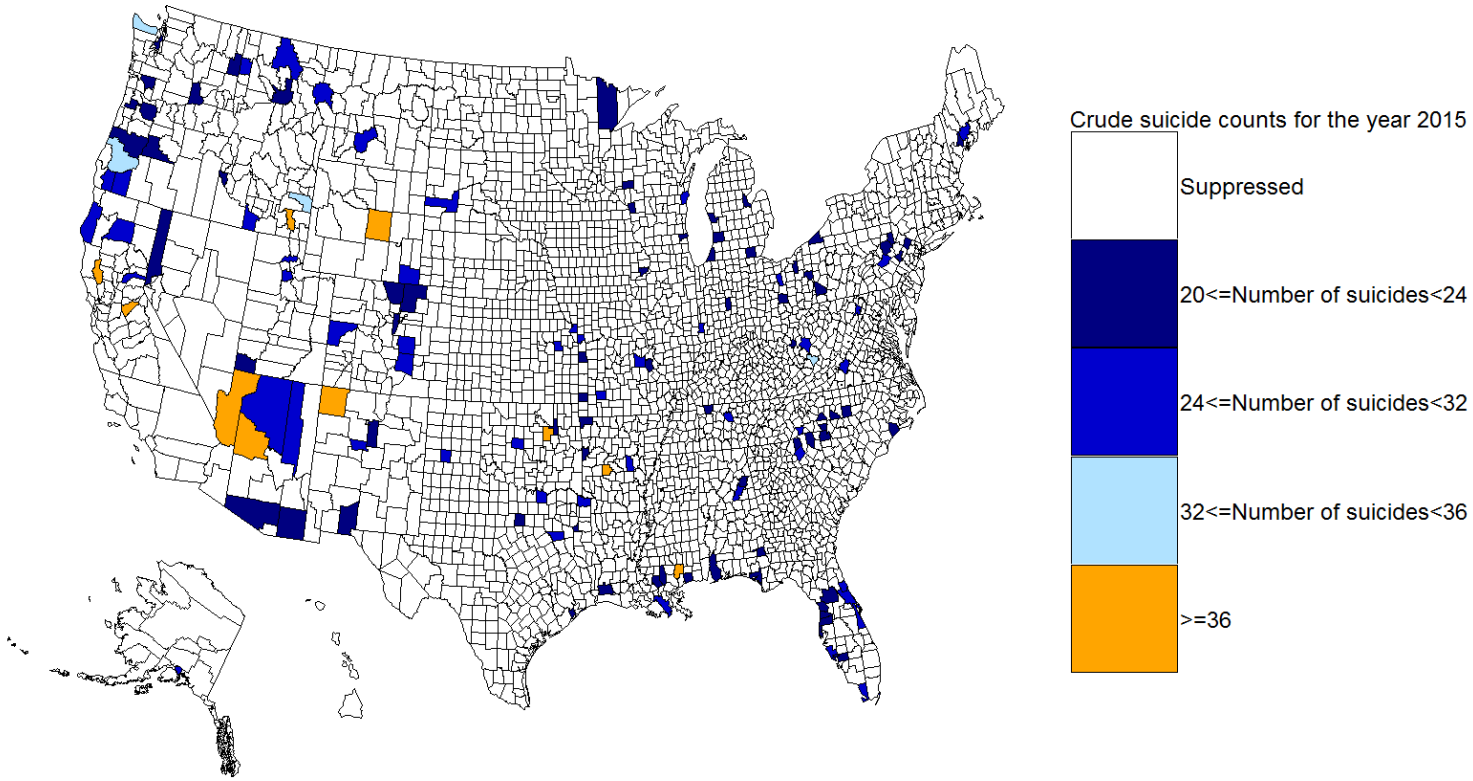


Figure: Suicide rates based on less than 20 suicides for a county are suppressed for the year 2015.

Why Bayesian approach?

1. Main advantage: takes into account the uncertainty of the estimates/predictions
2. The inferential process accounts for spatial trend via spatially structured random effects: providing insight knowledge
3. Hierarchal structure accounts for similarities based on neighbourhood structure
4. The space-time effects explain the differences/changes in time trend for different counties

Problem formulation

y_{it} : counts of suicides by county i and year t
n_{it} : counts of population by county i and year t
$y_{it} \sim Binomial(n_{it}, p_{it})$
p_{it} : probability of suicides in county i at time t
where $i = 1, \ldots, m$ areas and $t = 1, \ldots, T$ times.
General space-time model structure (Lawson, A. (2013)):
$logit(p_{it}) = \alpha_0 + A_i + B_t + C_{it}$,
where:
A_i : spatial group
B_t : temporal group
C_{it} : space-time interaction group

Proposed models

Model 1 With uncorrelated heterogeniety (non-spatial random effect)
$logit(p_{it}) = \alpha_0 + v_i$
Prior for $\alpha_0 \sim dflat$
Prior for v_i: $v_i \sim N(0, 1/\tau_v)$ termed as uncorrelated heterogeneity (variability)
τ_v is the precision
$Log(\tau_v) \sim LogGamma(1, 0.001)$

Proposed models

Model 2 Besag: convolution model
$logit(p_{it}) = \alpha_0 + u_i + v_i$
Intrinsic CAR prior for:$u_i|u_{-i}$
$u_i|u_{-i} \sim N(\bar{u}_i, r/n_{\delta_i})$ termed as correlated heterogeneity (variability), where
$u_{-i} = (u_1, u_2, \ldots, u_{i-1}, u_{i+1}, \ldots, u_m)$
n_{δ_i} : number of neighbors $= \sum_{j=1}^{m} \omega_{ij}$
δ_i : neighborhood of ith region
r : is the variance, $r = 1/\tau_r$:τ_r is the precision
$Log(\tau_r) \sim LogGamma(1, 0.001)$
$\bar{u}_i = \frac{\sum_{j=1}^{m} \omega_{ij} u_j}{\sum_{j=1}^{m} \omega_{ij}}$
$\alpha_0 \sim dflat$

Proposed models

Model 3 Random time effect (correlated)
$logit(p_{it}) = \alpha_0 + u_i + v_i + \gamma 1_t$
$\gamma 1_t \sim N(\gamma 1_{t-1}, 1/\tau_{\gamma 1})$ (randomwalk)
$Log(\tau_{\gamma 1}) \sim LogGamma(1, 0.001)$

Proposed models

Model 4 Random time effect (uncorrelated)
$logit(p_{it}) = \alpha_0 + u_i + v_i + \gamma 2_t$
$\gamma 2_t \sim N(0, 1/\tau_{\gamma 2})$ (uncorrelated)
$Log(\tau_{\gamma 2}) \sim LogGamma(1, 0.001)$

Proposed models

Model 5 Space-time interaction (uncorrelated)
$logit(p_{it}) = \alpha_0 + u_i + v_i + \gamma 1_t + \psi_{it}$
$\psi_{i,t} \sim N(0, 1/\tau_\psi)$ (uncorrelated)
$Log(\tau_\psi) \sim LogGamma(1, 0.001)$

Covariates

1. Covariates enhance small area predictions
2. 53 variables identified based on past literature studies
3. Demographic variables, health-related characteristics, socioeconomic factors, treatment gap for drug and alcohol use, county level model based estimates of drug poisoning

Proposed models

Model 6 Model 5 plus covariates
$logit(p_{it}) = \alpha_0 + u_i + v_i + \gamma 1_t + \psi_{it} + \mathbf{X}'\boldsymbol{\beta}$
$\psi_{i,t} \sim N(0, \tau_\psi)$ (uncorrelated)
$Log(\tau_\psi) \sim LogGamma(1, 0.001)$
$\mathbf{X}$: time varying covariates matrix X_{it}
$\boldsymbol{\beta}$: vector of regression parameters
$\beta \sim N(0, 100)$

Goodness of the fit: Deviance Information Criterion (DIC) and Watanabe-Akaike information criteria (WAIC)

Table: DIC, the effective number of parameters estimated and WAIC for the models via INLA.

Model	DIC	n.eff	WAIC
Model 1	150371.4	2316	150763.3
Model 2	149966.2	2316	150280.2
Model 3	148008.6	1884	148166.2
Model 4	148010.3	1886	148168.1
Model 5	147821.9	2766	147938
Model 5 + covs	**147181.1**	**1896**	**147250**

Model accuracy: State direct Suicide Rates vs INLA model based aggregated state level Suicide Rates, 2005-2015

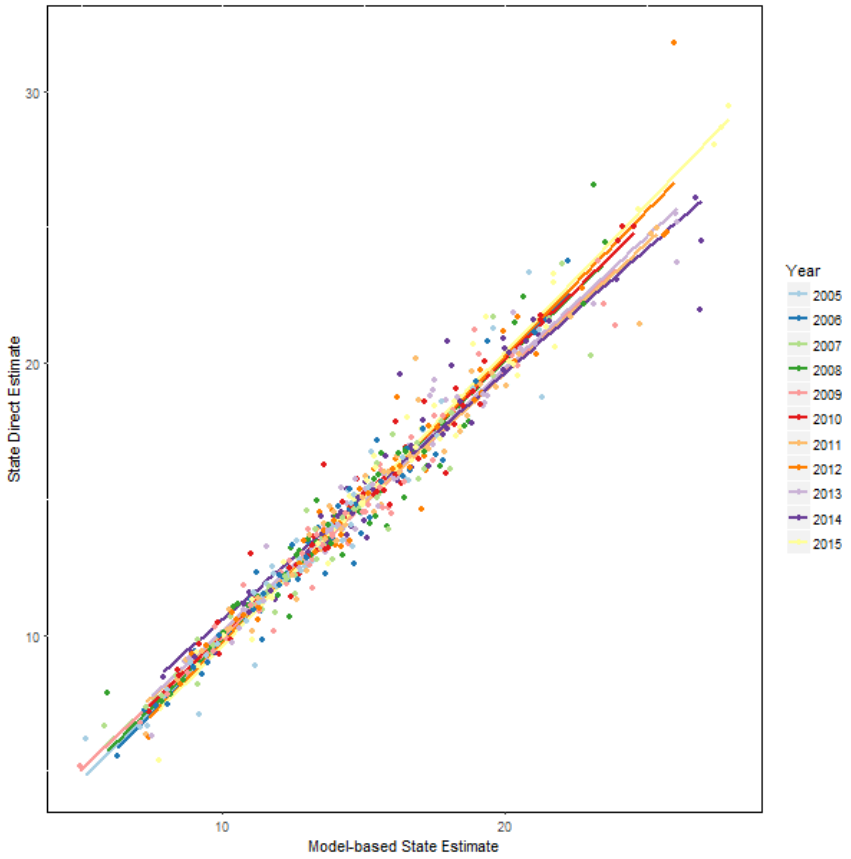


Figure: State direct SRs vs INLA based state SRs 2005-2015.

Shrinkage of model based Suicide Rates for each state, by population size for 2015

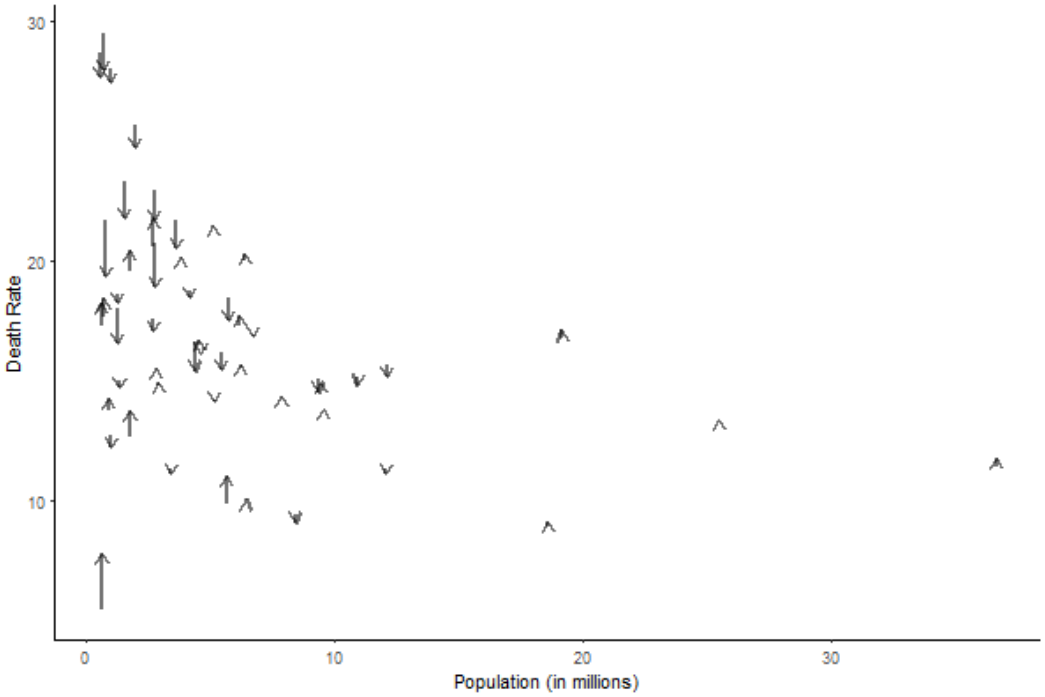


Figure: Shrinkage of suicide rates for each state, by population size for 2015. Crude death rates are plotted at the start of the arrows, and model-based death rates are located at the end of the arrows.

Model accuracy: MCMC SRs vs INLA SRs, **Computation time: MCMC: (8 weeks) vs INLA (24-36 hours)**

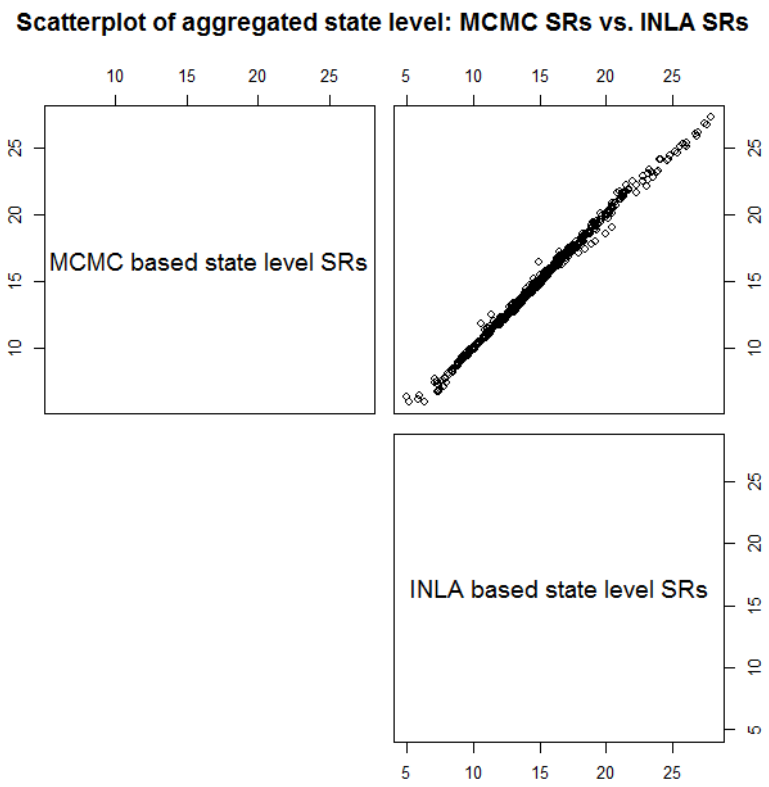


Figure: MCMC based vs INLA based aggregated state level Suicide Rates, 2005-2015.

Predicted Suicide Rates (SRs) 2005

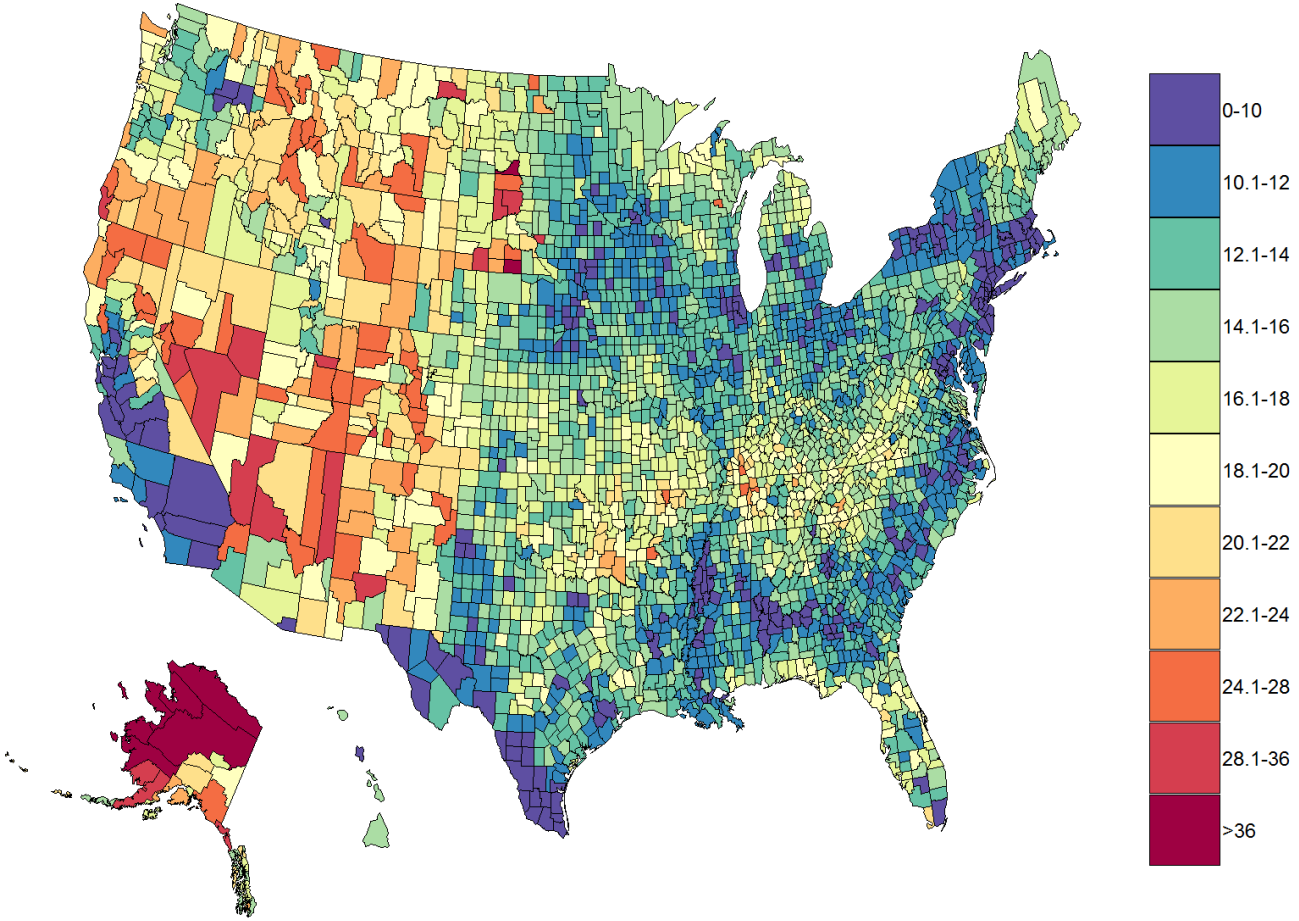


Figure: Predicted posteriors 2005.

Predicted Suicide Rates (SRs) 2015

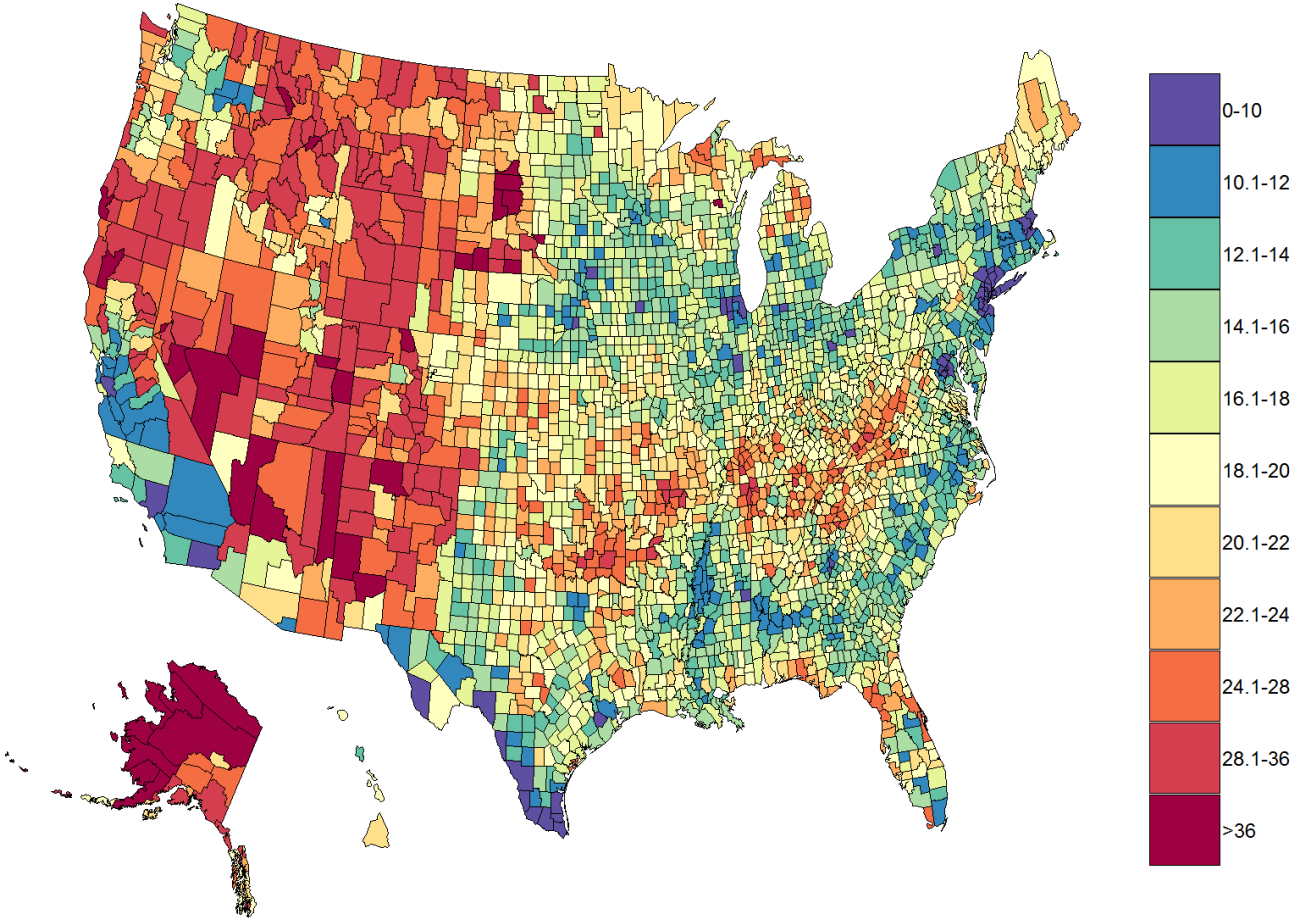
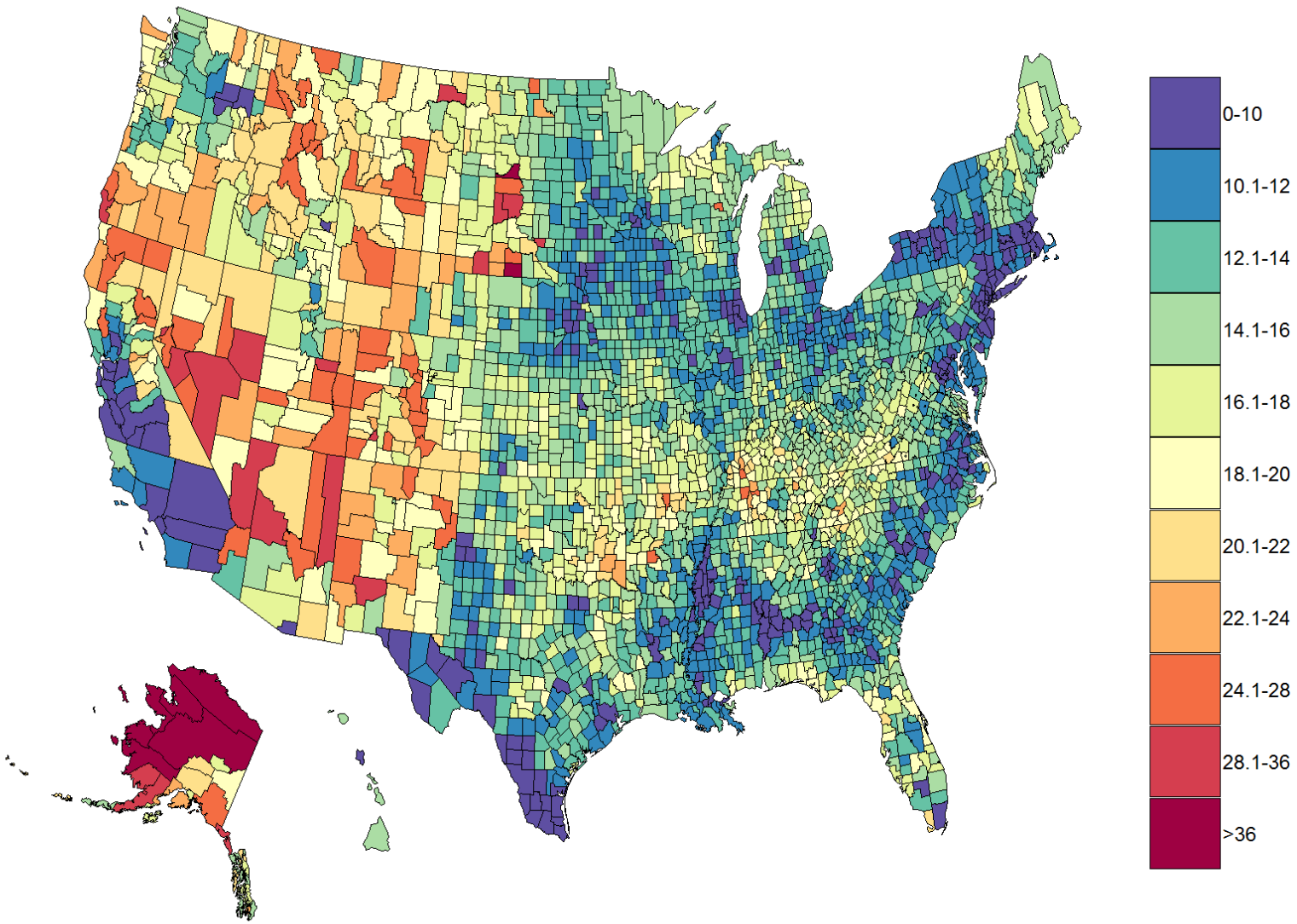


Figure: Predicted posteriors 2015.

Animation of Suicide Rates: 2005 through 2015



Covariates significantly associated with SRs

1. Demographic characteristics: household size, racial and ethnic distribution, urbanization level, and divorce rates
2. Socioeconomic factors: median home value, median gross rent, household crowding, and median per capita income, percent persons with college education, unemployment rate, high-cost loan rate
3. Health-related characteristics: percent abusing or dependent on illicit drugs or alcohol in the previous year, treatment gap for alcohol and drug use, and prevalence of major depressive episode
4. County-level model-based estimates of age-adjusted death rates due to drug poisoning
5. Consistent with prior analyses reporting county-level (i.e., ecological) associations between socioeconomic, demographic and/or health-related factors and suicide rates

Exceedance probabilities

1. The probability that the predicted SRs exceed a certain threshold
2. **Can be used to quantify the uncertainty associated with the county level SRs. Example: Probability that the predictions exceed the mean county level SRs**

Uncertainty associated with the county level SRs, year 2005

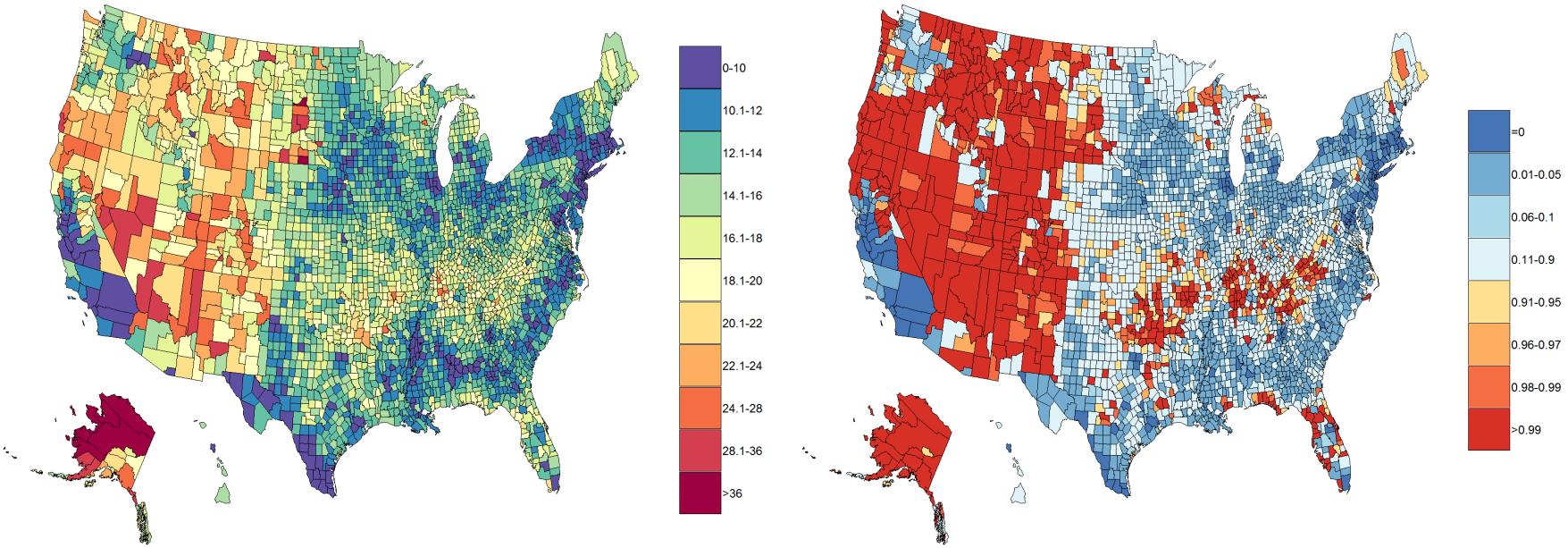


Figure: Predicted county level SRs and the probability of exceeding the crude mean county level SR for the year 2005 (14.61 per 100,000).

Uncertainty associated with the county level SRs, year 2015

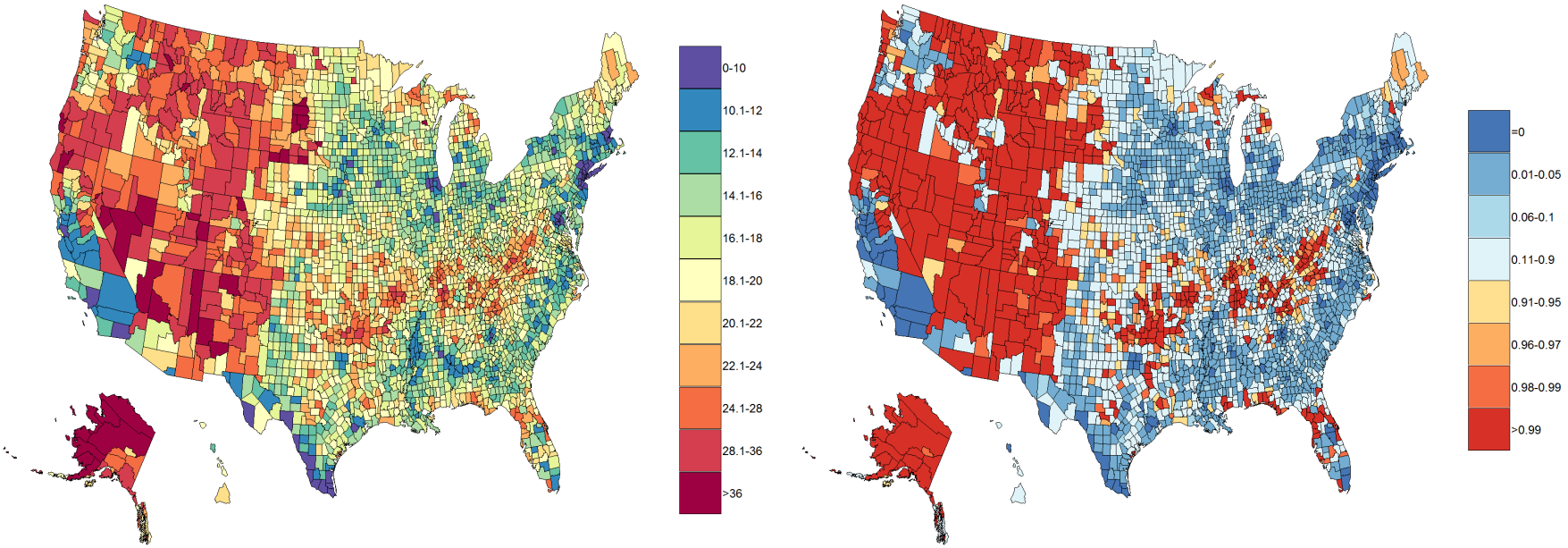


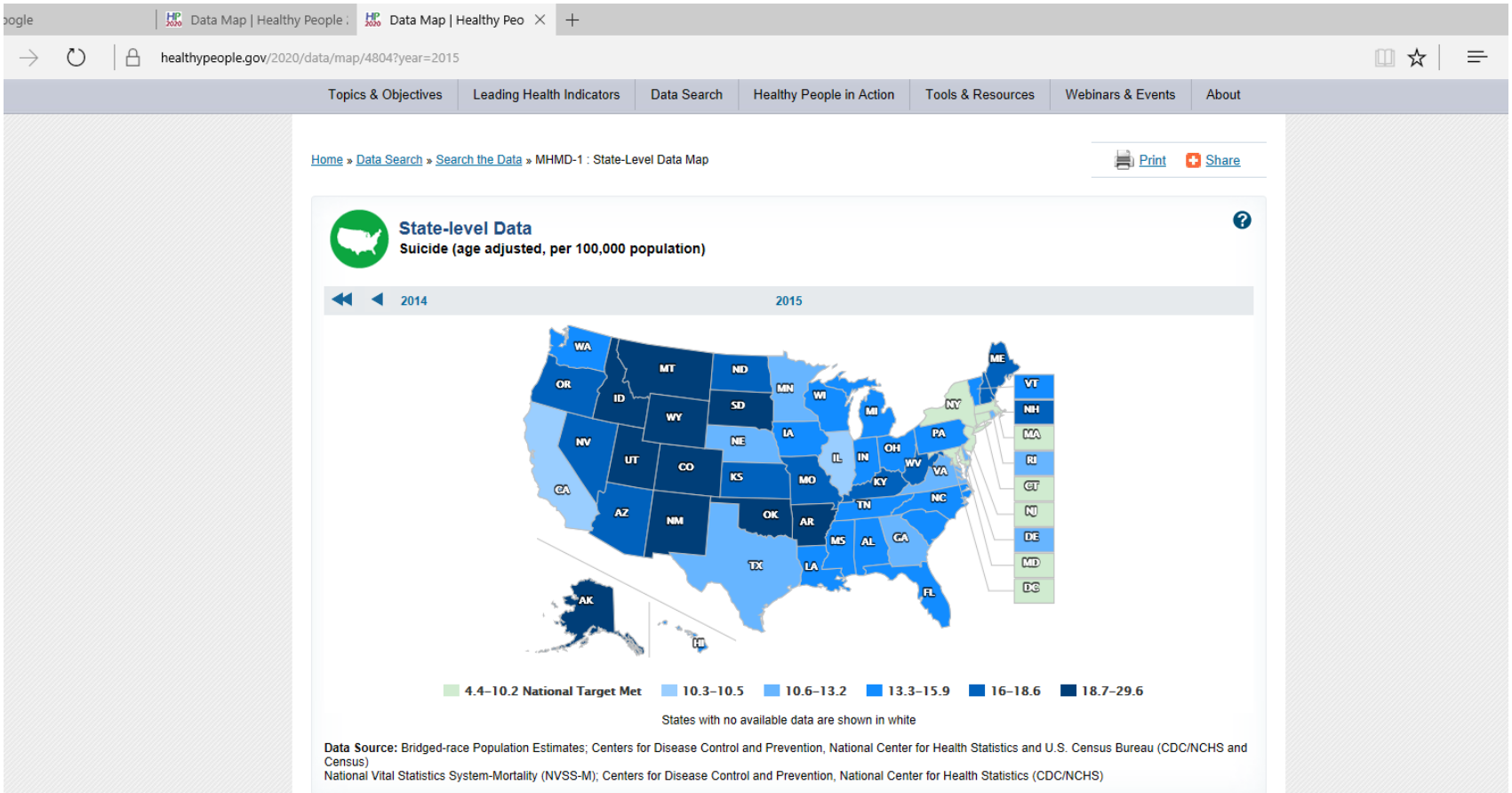
Figure: Predicted county level SRs and the probability of exceeding the crude mean county level SR for the year 2015 (18.74 per 100,000).

Exceedance probabilities

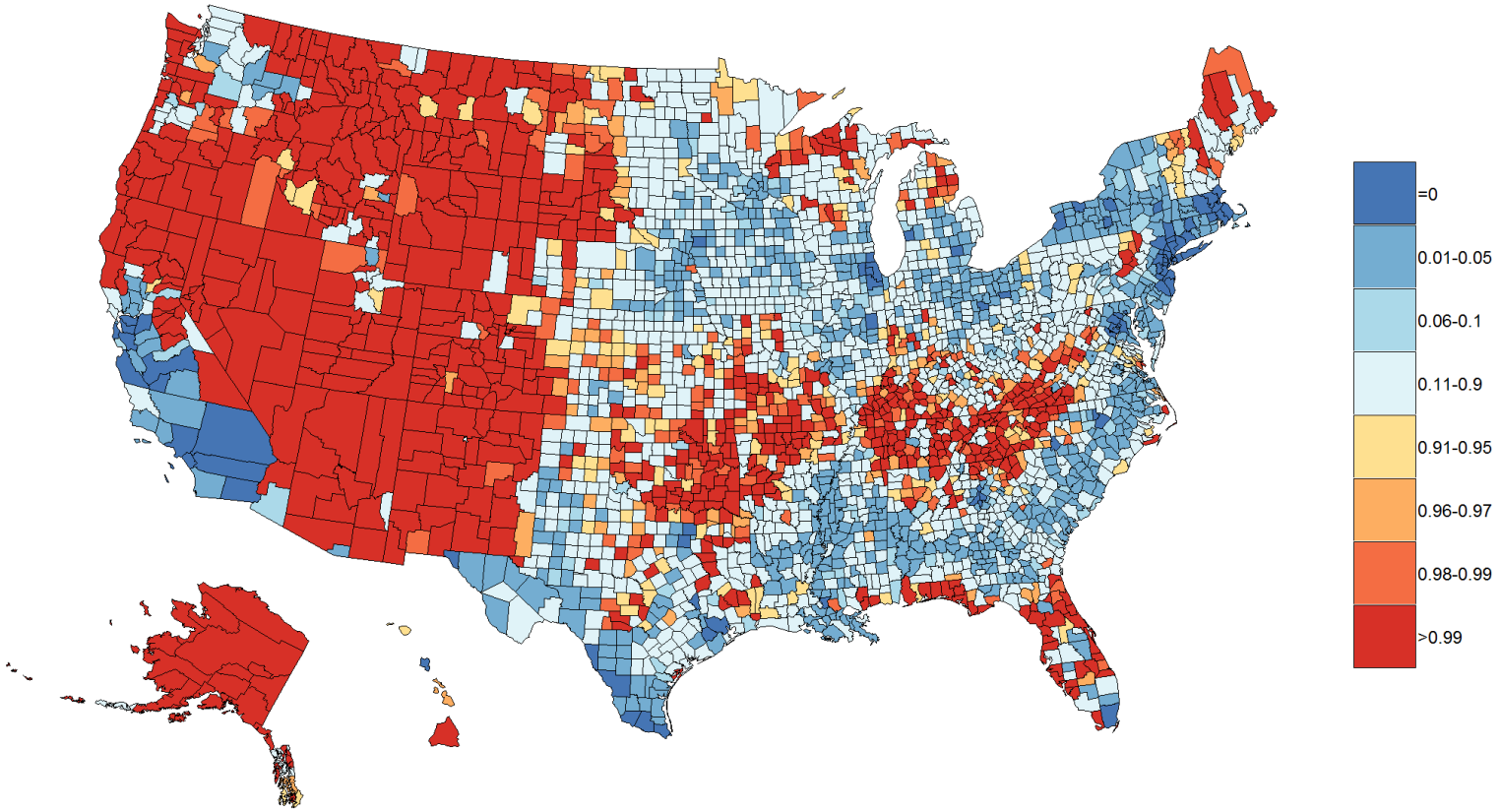
1. Healthy People 2020 (HP2020) sets targets
2. HP2020 uses the age-adjusted suicide rate for 2007 for state level SRs, which is 11.3 per 100,000 to set targets. Apply a 10 percent improvement to get the target of 10.2 per 100,000
3. For crude county level suicide rates: mean crude rate for 2007 is 14.91 per 100,000. Apply a 10 percent improvement to get the **target** of 13.419 per 100,000

State-level Data 2015: Suicide rate (age adjusted, per 100,000 population)

https://www.healthypeople.gov/2020/data/map/4804?year=2015



Pr (exceeding the target rate of 0.00013419) - 2005-2015.



Trends in suicide rates: correlated time effect, Type II random walk, $\gamma 2_t$ and 95 percent Bayesian credible intervals

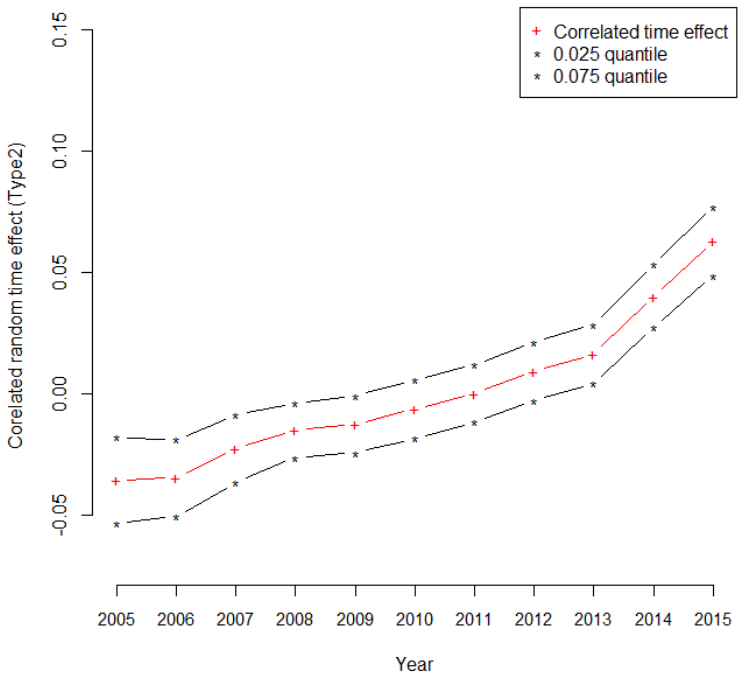


Figure: Correlated time effect $\gamma 2_t$.

Uncorrelated heterogeneity and correlated heterogeneity

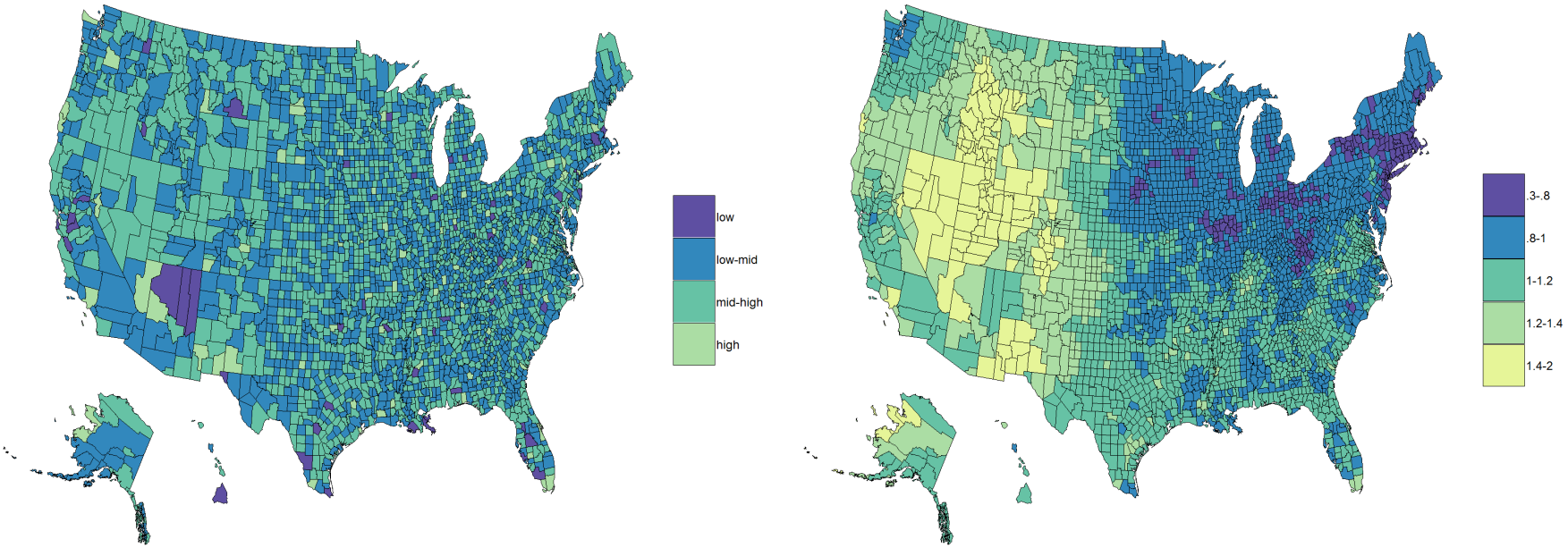


Figure: Uncorrelated heterogeneity v_i and correlated heterogeneity u_i.

Conclusions

1. Use of R-INLA method resulted in **substantially reduced computation time: 8 weeks vs 24 hours**
2. A variety of time and random effects could be tested
3. **R-INLA Allows faster computation of exceedance probabilities to determine if counties have met the specified targets/thresholds**
4. The functional form of the covariates in R-INLA can be specified in different forms

Conclusions

1. TBRs declined across all regions of the country from 2003 to 2015
2. TBRs remained in excess of 67 births per 1,000 adolescent females in several counties across Texas, along the Mississippi river, Montana, New Mexico, Georgia and Alaska
3. Higher TBRs across counties in the southern U.S. and lower TBRs in New England counties during the study period, 2003 to 2015
4. Large variation in TBRs in smaller counties within states and large teen birth rates in rural areas
5. Data visualization: can zoom in on counties for more granular look
6. Data visualization: identify counties in greatest need

Conclusions

1. All counties demonstrated an increase in suicide death rates from 2005-2015
2. The counties with the highest suicide rates were predominantly located across the western US
3. The counties with lowest rates were observed across southern California, western Texas, along the Mississippi river, and in areas along the East Coast
4. Several county-level covariates, namely socioeconomic, demographic,and/or health related factors were found to be significant predictors of SRs

Future research

1. Smoothed county level estimates can be used to assess urban-rural disparities
2. Future research examining spatiotemporal trends by age and gender would be informative
3. Spatial clustering at the sub county levels would provide additional insights
4. Future research can look at county-level variation by race and Hispanic origin groups
5. Future research on Neonatal intensive care units (NICU) births at the county level would be informative

Ongoing research: Hot and cold spots for Teen Birth Rates, females aged 15-19, 2016 for Hispanics

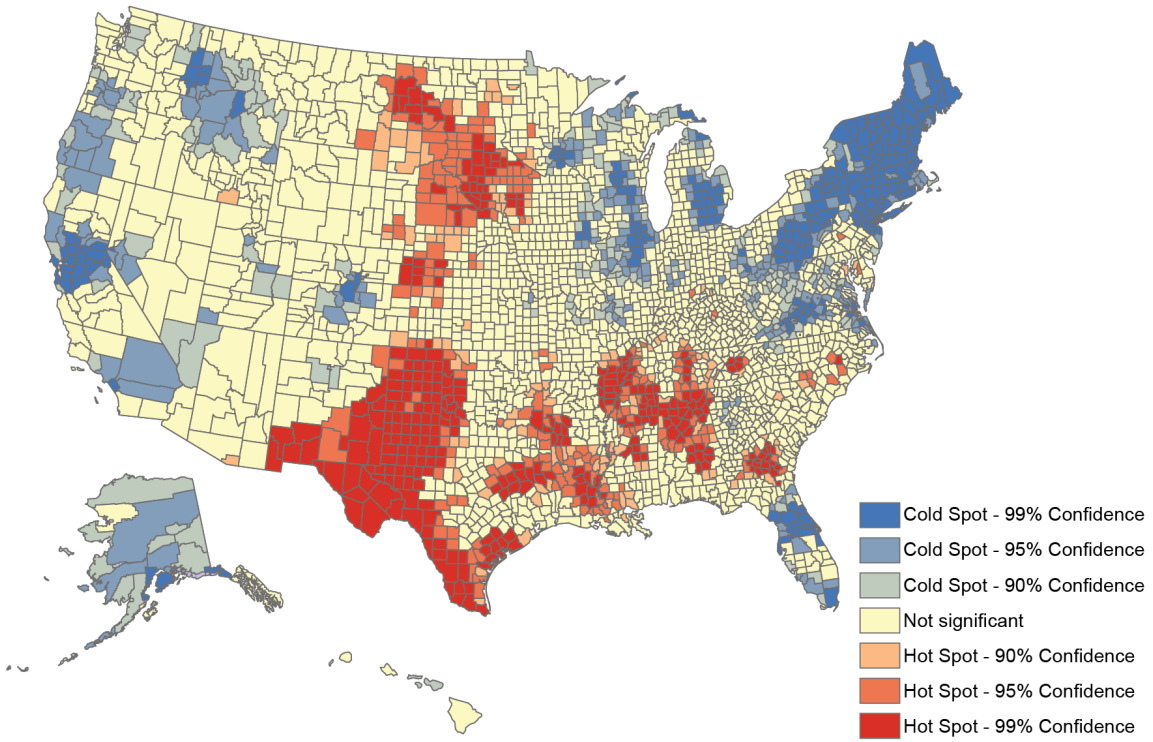


Figure: Values represent z-scores from the Getis Ord Gi analysis via ArcGis. Negative z-scores indicate cold spots, while positive z-scores indicate hot spots.

References

1. Khan, D., Rossen, L.M., Hamilton, B., Dienes, E., He, Y. and Wei, R. (2017). Spatiotemporal trends in teen birth rates in the USA, 20032012. J. R. Stat. Soc. A. http://onlinelibrary.wiley.com/doi/10.1111/rssa.12266/epdf
2. Khan, D., Rossen, L.M., Hedegaard, H., Warner, M. A Bayesian Spatial and Temporal modeling approach to Mapping Geographic Variation in Mortality Rates for Subnational Areas with R-INLA. Journal of Data Science. 18(2018), 147-182. http://jds-online.com/
3. Martin, J.A., Hamilton, B.E., Osterman, M.J.K., et al. Births: Final data for 2015. National vital statistics reports; vol 66 no 1. Hyattsville, MD: National Center for Health Statistics. 2017.

References

1. Khan, D., Rossen, L. M., Hamilton, B.E., Dienes E, He, Y., Wei, R. Hot Spots, Cluster Detection and Spatial Outliers Analysis of Teen Birth Rates in the U.S., 2003 to 2012. Spatial and Spatio-temporal Epidemiology, 21:67-75, June 2017. www.sciencedirect.com/science/article/pii/S1877584516300442
2. Hamilton, B.E. et al. U.S. and state trends on teen births, 19902015. National Center for Health Statistics. 2017. https://wwwdev.cdc.gov/nchs/data-visualization/teen-births/

References

1. Rue, H. and Martino, S. INLA: Functions which allow to perform a full Bayesian analysis of structured additive models using Integrated Nested Laplace Approximation. 2009. R package, version 0.0.
2. Rue, H., Martino, S., and Chopin, N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. Journal of the Royal Statistical Society Series B-Statistical Methodology, 71:319-392. 2009.
3. http://www.r-inla.org/
4. Rossen, L.M., Hedegaard, H., Khan, D., and Warner, M. County level Trends in Suicides Rates in the U.S., 2005-2015. Accepted for publication in AJPM.

Contact information

1. Diba Khan *ild1@cdc.gov* or 301-458-4474
2. For teen births visualization: Brady E. Hamilton *boh5@cdc.gov* or 301-458-4653