

Supplementary Information: A data-driven semiparametric model of SARS-CoV-2 transmission dynamics in the United States

2022-08-03

John M. Drake, Andreas Handel, Éric Marty, Eamon B. O'Dea, Tierney O'Sullivan, Giovanni Righi, and Andrew T. Tredennick,

Model diagram

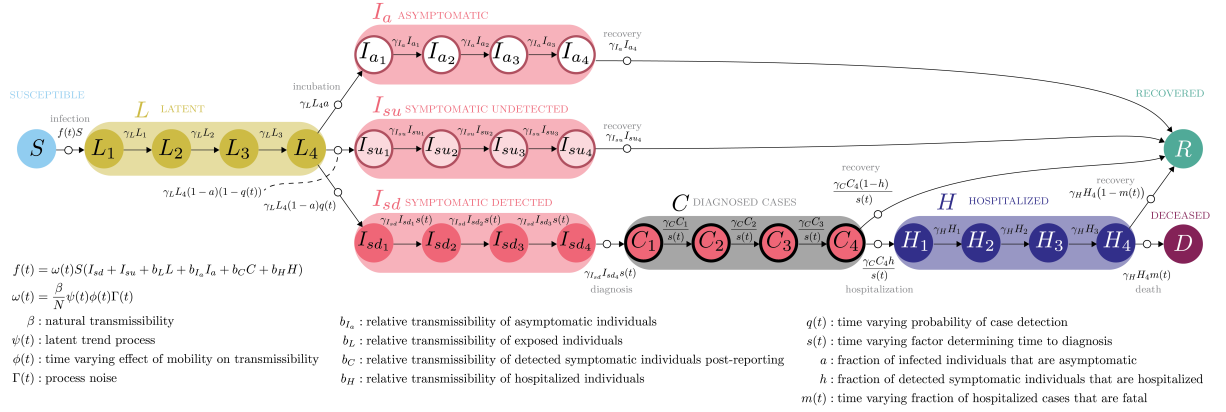


Figure S1: Model diagram.

Model parameters

Table S1: Fixed model parameters.

Parameter definition	Parameter symbol	Value	Source
Reproduction number	\mathcal{R}_0	7	Sanche et al. (2020)
Baseline transmission rate of symptomatic individuals	β	$\mathcal{R}_0 \times 0.1 \times \frac{1}{N}$	Assumption
Fraction of infected individuals that are asymptomatic	a	0.18	Mizumoto et al. (2020);
Fraction of diagnosed cases that are hospitalized	h	0.12	Verity et al. (2020)
Relative transmissibility of L to $I_{su/sd}$	b_L	0.12	Li et al. (2020)
Relative transmissibility of I_a to $I_{su/sd}$	b_{I_a}	0.5	Li et al. (2020)
Relative transmissibility of C to $I_{su/sd}$	b_C	0.27	Assumption
Relative transmissibility of H to $I_{su/sd}$	b_H	4.5×10^{-5}	Moghadas et al. (2020)
Duration of time in L stages	$1/\gamma_L$	4 days	Li et al. (2020)
Duration of time in I_a stage	$1/\gamma_{I_a}$	3.5 days	Li et al. (2020)
Duration of time in I_{su} stages	$1/\gamma_{I_{su}}$	6 days	Li et al. (2020)
Duration of time in I_{sd} stage	$1/\gamma_{I_{sd}}$	$0.5 \times 1/\gamma_{I_{su}}$	Assumption
Duration of time in C stages	$1/\gamma_C$	$0.5 \times 1/\gamma_{I_{su}}$	Assumption
Duration of time in H stages	$1/\gamma_H$	6 days	Ohfeldt et al. (2020)
Minimum detection probability	q_{\min}	0.1	Assumption
Maximum detection probability	q_{\max}	0.4	Assumption
Day at which detection probability is halfway between q_{\min} and q_{\max}	q_{half}	30	Assumption
Rate of increase from q_{\min} to q_{\max}	q_r	1.0	Assumption
Maximum factor by which diagnosis speed increases	s_{\max}	1.0	Zhang et al. (2020)
Day at which diagnosis speed-up factor (s) is halfway between 0 and s_{\max}	s_{half}	30	Assumption
Rate of increase from 0 to s_{\max}	s_r	1.0	Assumption
Initial size of susceptible pool	$S(t=1)$	N_{state}	Assumption

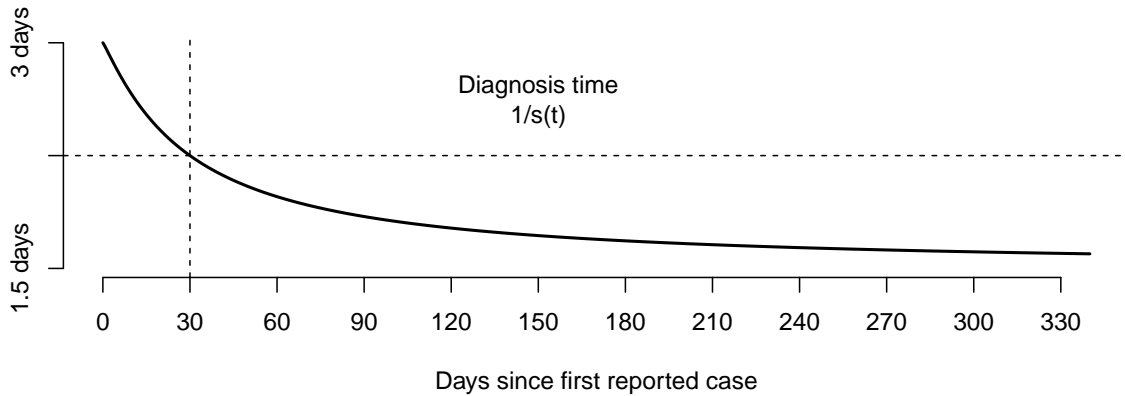
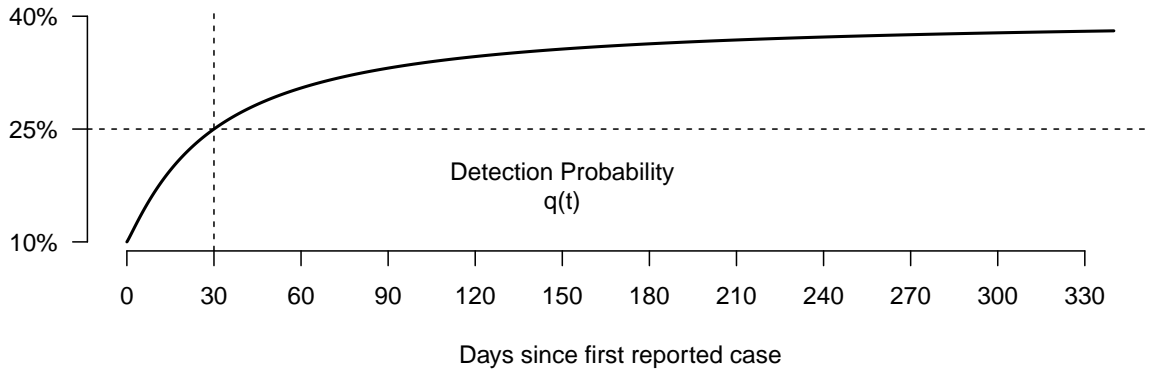
Table S2: Estimated parameters and starting ranges for MIF estimation procedure. The “expit” function refers to back-transforming the parameter from the logit scale, which was used for estimation.

Parameter definition	Parameter symbol	Start range
Baseline fraction of hospitalizations that result in death	m_{base}	[expit(-6), expit(6)]
Minimum fraction of hospitalizations that result in death	m_{min}	[expit(-6), expit(6)]
Day at which death fraction is halfway between m_{base} and m_{min}	m_{half}	[exp(-5), exp(5)]
Extra-demographic process noise	σ	[exp(-5), exp(5)]
Case reporting dispersion	θ_c	[exp(-5), exp(5)]
Death reporting dispersion	θ_d	[exp(-5), exp(5)]
Initial size of latent compartment	$L(t = 0)$	[exp(0), exp(10)]
Initial size of asymptomatic infectious compartment	$I_a(t = 0)$	[exp(0), exp(10)]
Initial size of undetected infectious compartment	$I_{su}(t = 0)$	[exp(0), exp(10)]
Initial size of detected infectious compartment	$I_{sd}(t = 0)$	[exp(0), exp(10)]
B-spline coefficients	g_i	[-10, 10]

Detection probability and diagnosis speed-up functions

The fraction of exposed individuals that are detected and flow into the I_{sd} compartments (q) starts at a low level of 0.1 and increases sigmoidally towards a maximum of 0.4, reaching the half way point of 0.25 on the 30th day since the first case notification. The time to diagnosis ($1/s$) decreases sigmoidally over time from 3 days on day 0 towards a minimum of 1.5 days, reaching the half way point of 2.25 days on the 30th day since the first case notification.

Both q , s follow a Hill function with Hill coefficient 1.1 and time to half maximum of 30 days:



Mean absolute scaled errors (MASE)

The MASE is a ratio of mean absolute error (MAE) in the model to MAE of an in-sample naïve model (a random walk forecast from the last observation, adjusted for weekly seasonality). MASE values were calculated for each of the 500 particle filter replicates for each state using the **yardstick** package in R. Most states had a mean MASE less than one for cases, while MASE was greater than one for deaths for most states. Thus, our model does not always outperform the benchmarking model, but it does offer mechanistic insight that the random walk model does not.

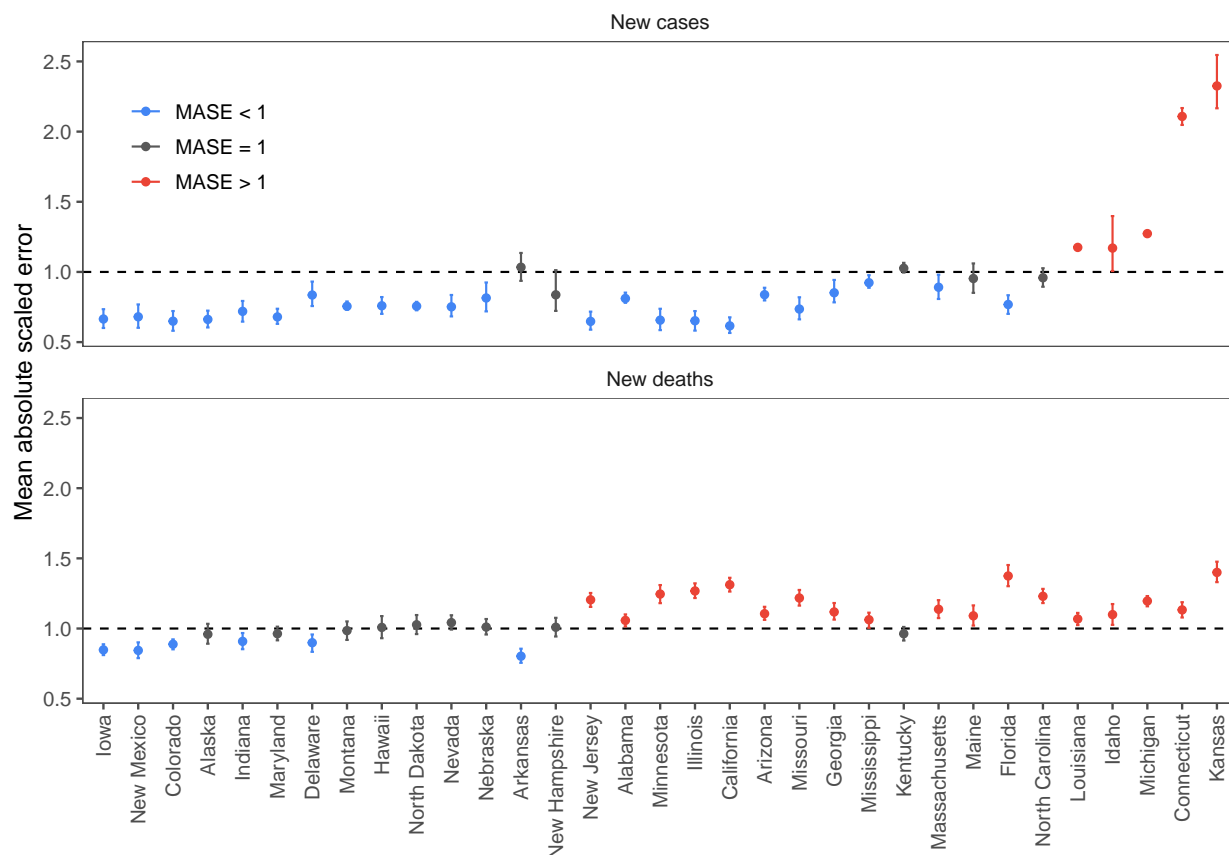
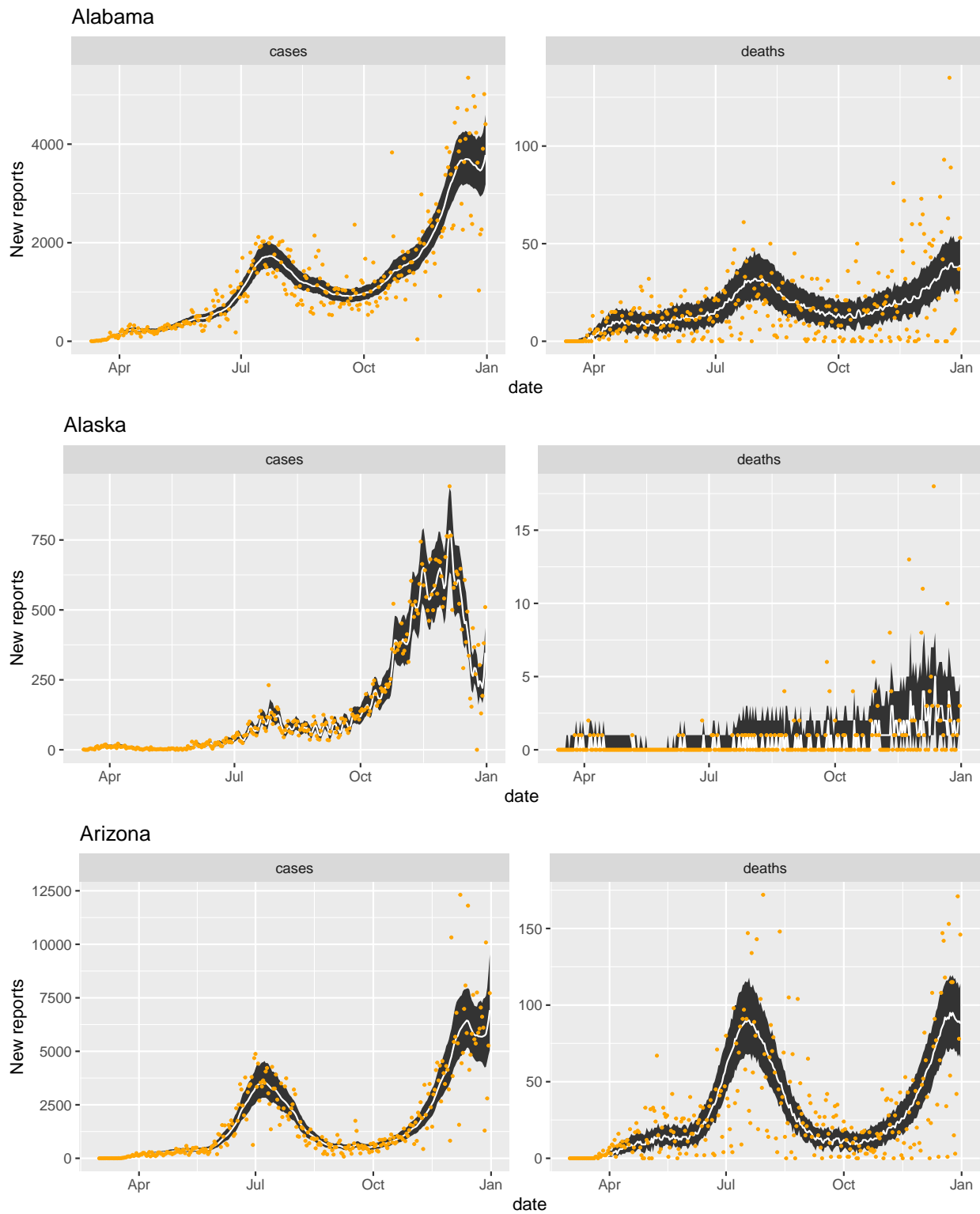


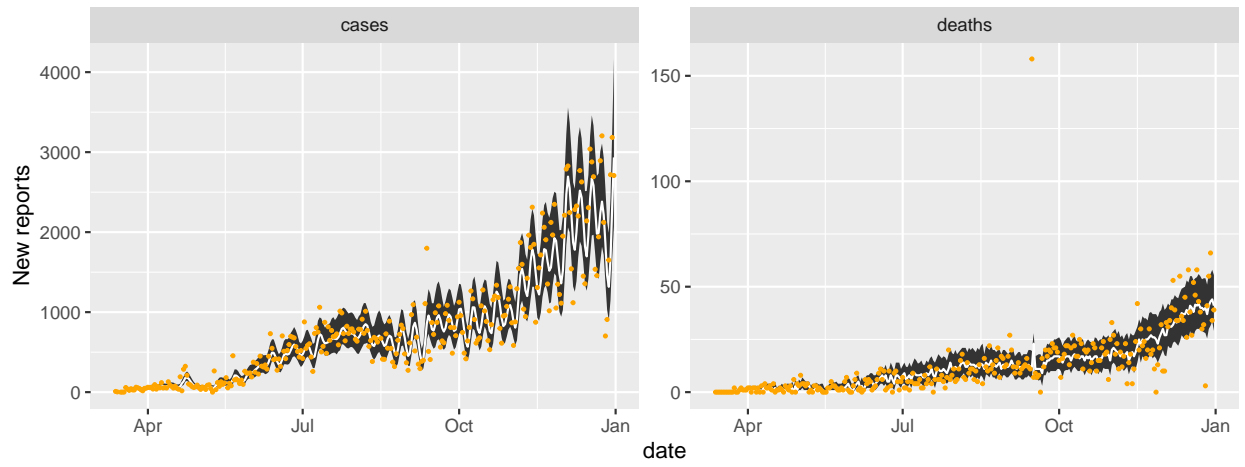
Figure S2: Mean and 95% intervals of mean absolute scaled errors (MASE) for each state. MASE was calculated for each of the 500 particle filter replicates for each state, yielding a distribution of MASE values. The point shows the mean of the MASE distribution and the errorbars show the 95% interval (lower 0.025 quantile to upper 0.975 quantile). States are ranked approximately from lowest MASE to highest MASE across both cases and deaths.

Time series of incident case and death reports

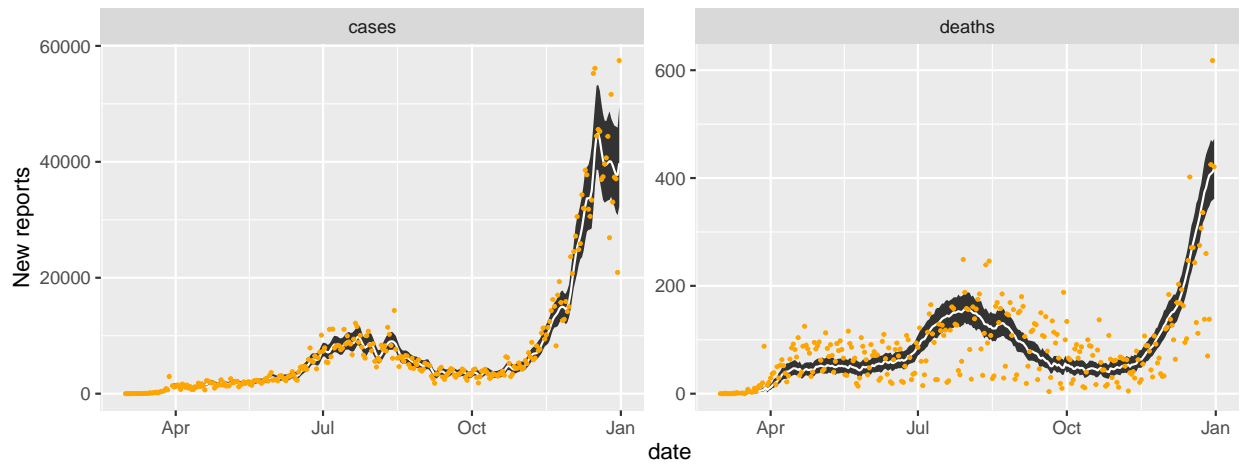
In the following plots, orange points are the reported daily cases or deaths, white line shows the median of the smoothed filtering distributions, and the grey ribbon bounds the 95% prediction interval.



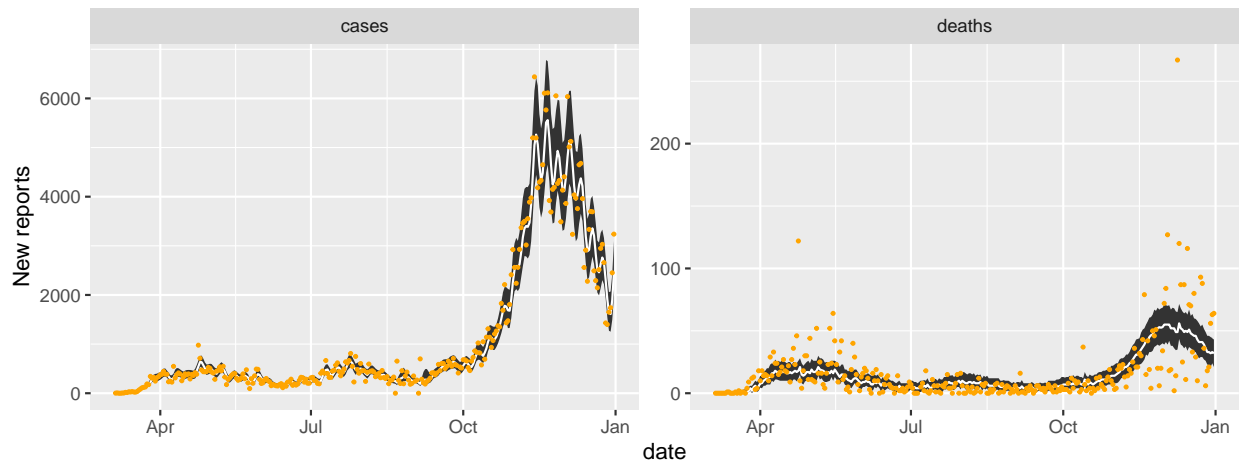
Arkansas



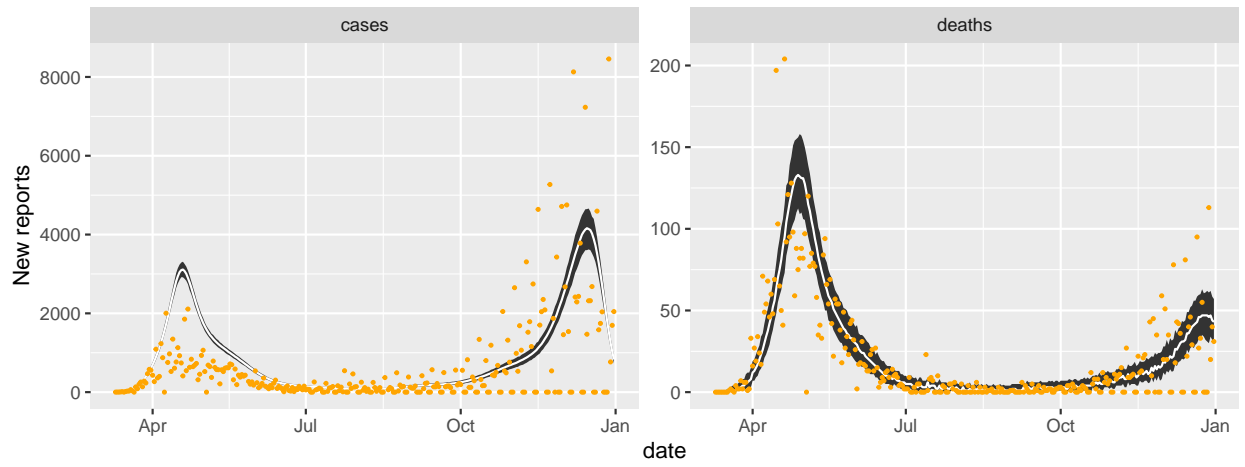
California



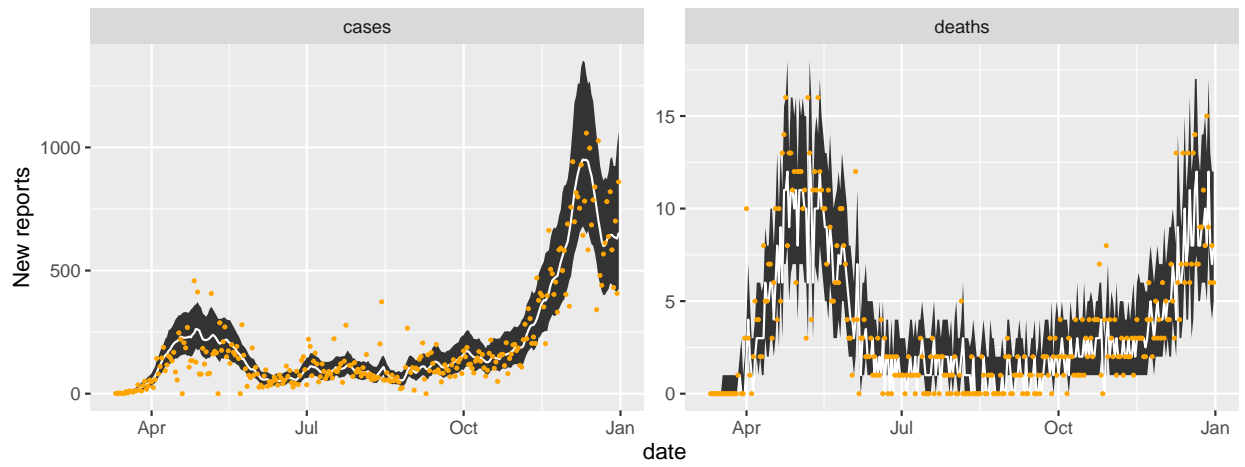
Colorado



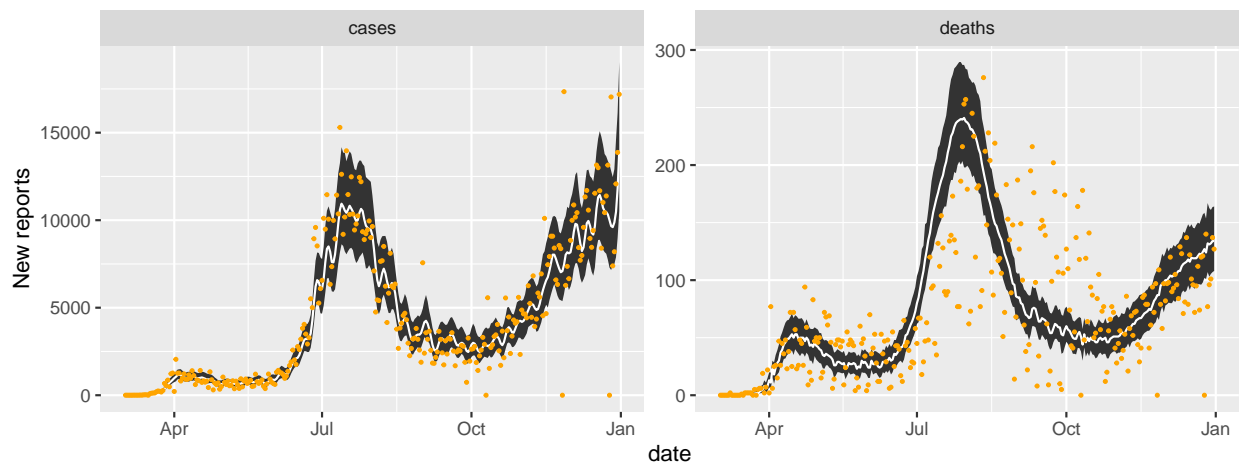
Connecticut



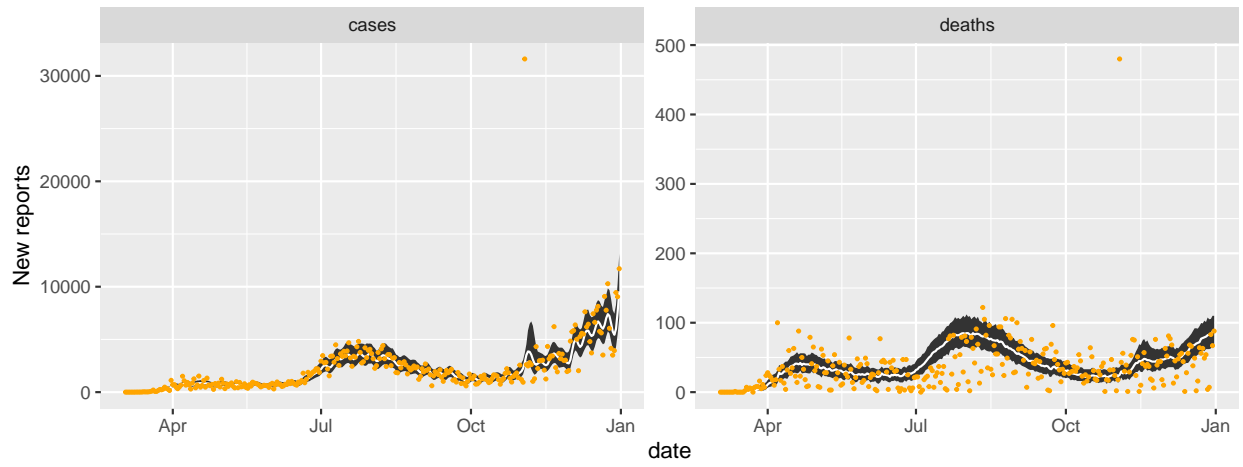
Delaware



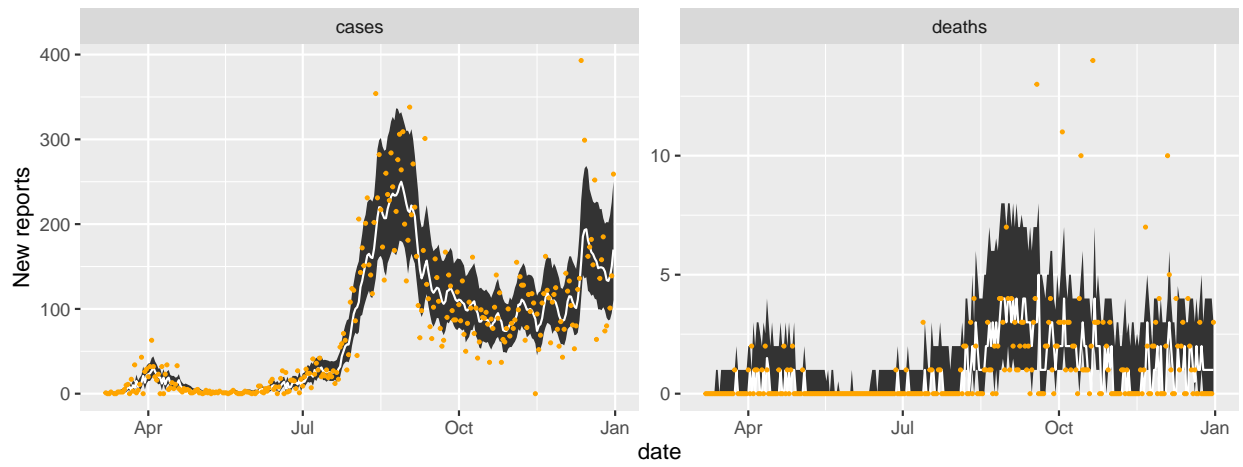
Florida



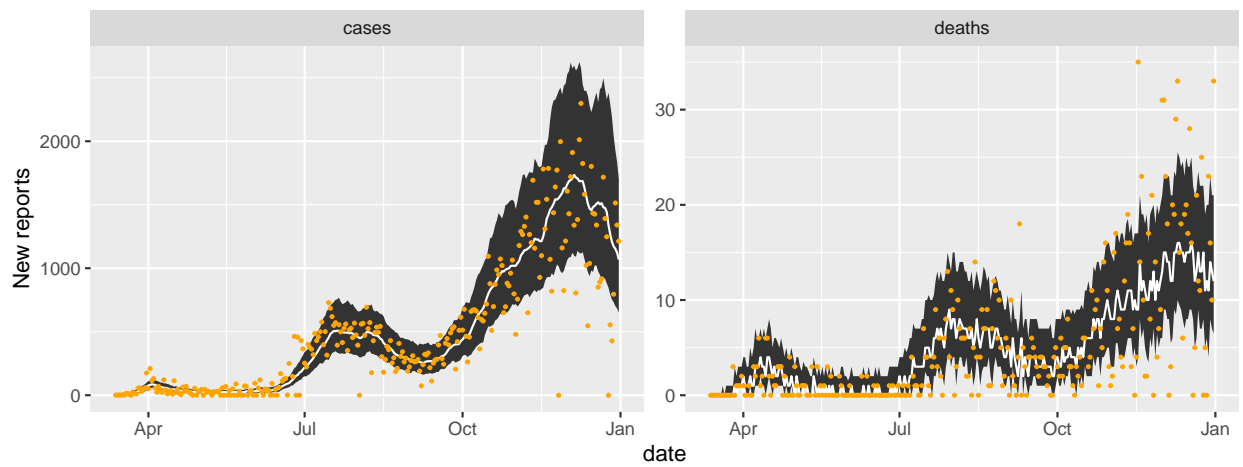
Georgia



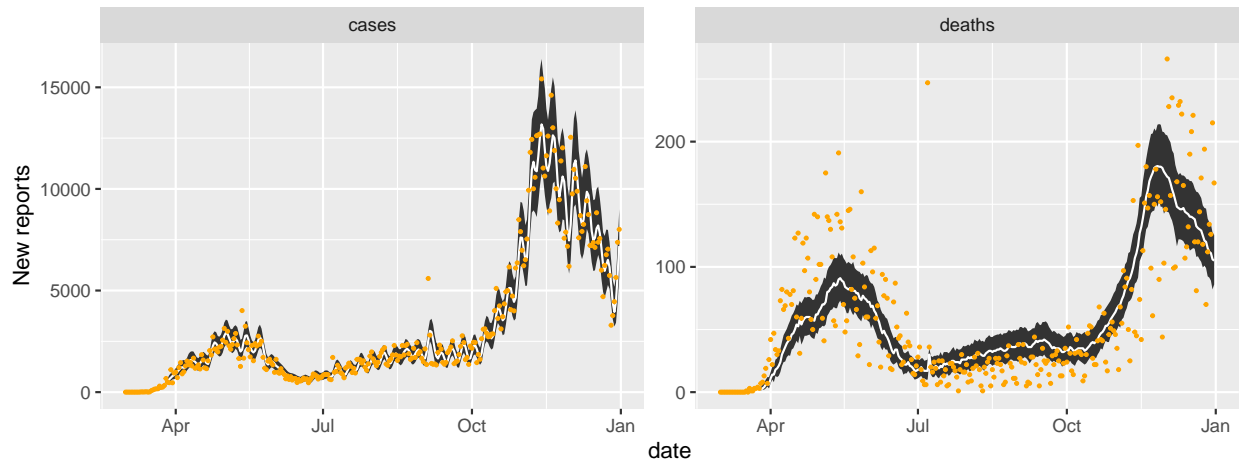
Hawaii



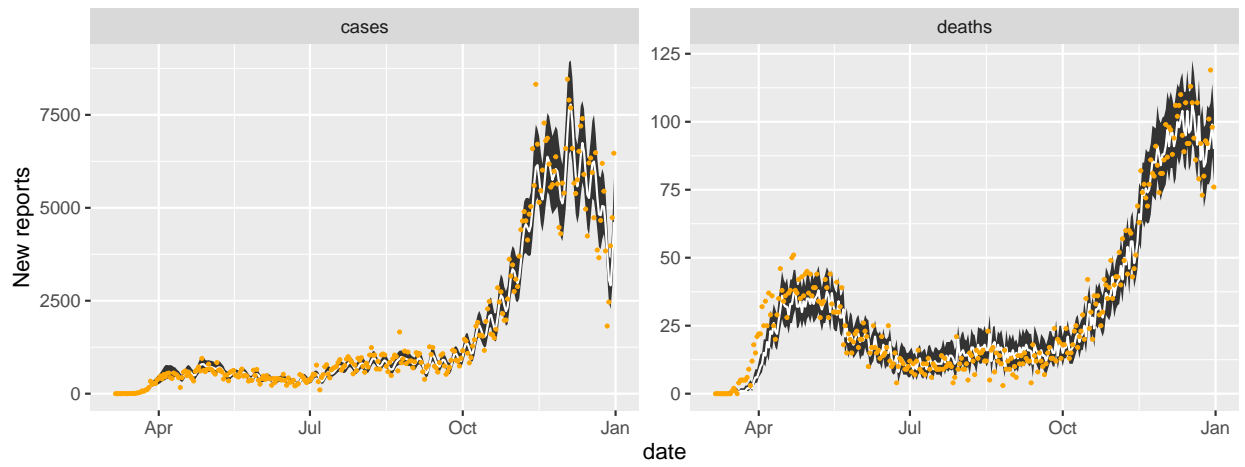
Idaho



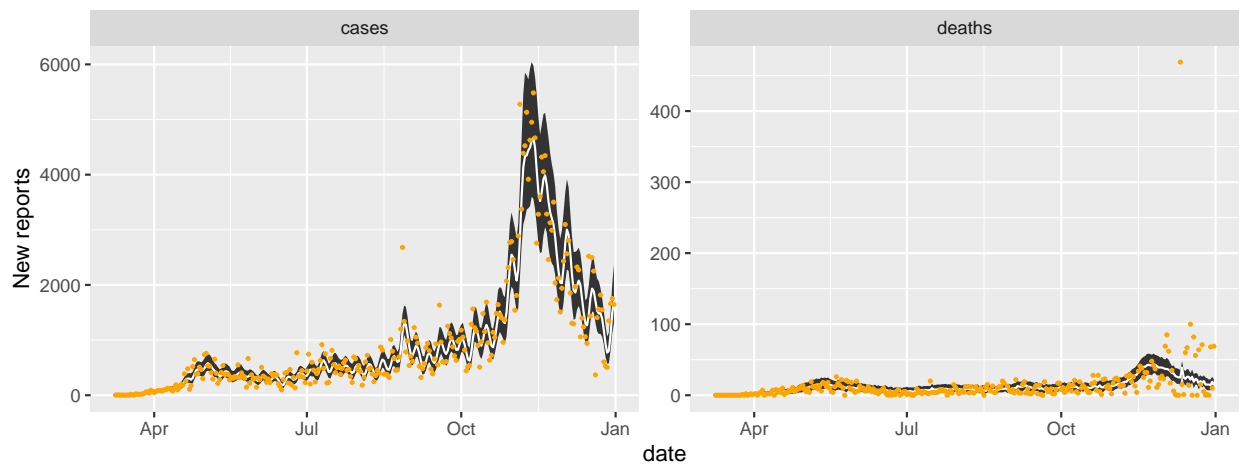
Illinois



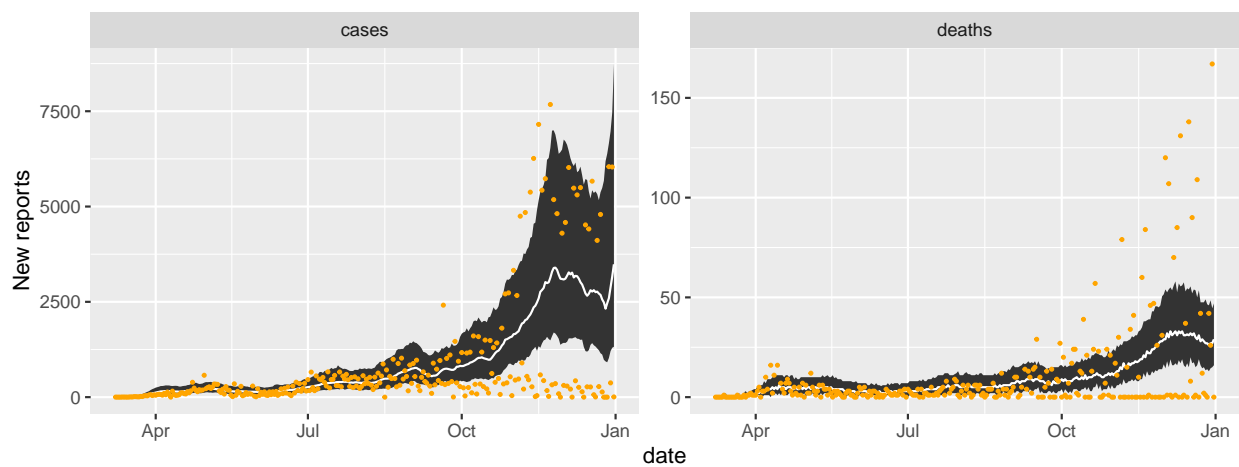
Indiana



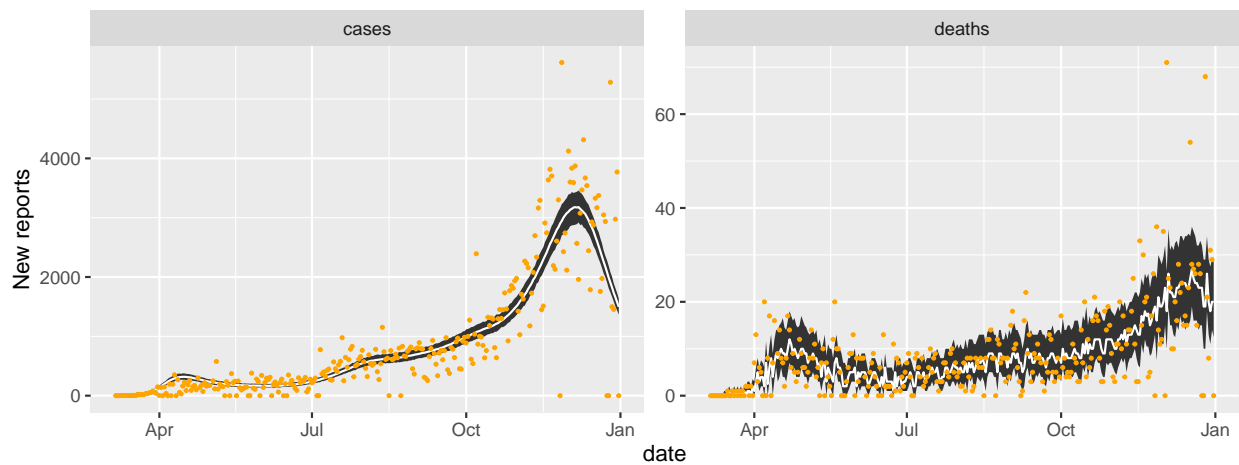
Iowa



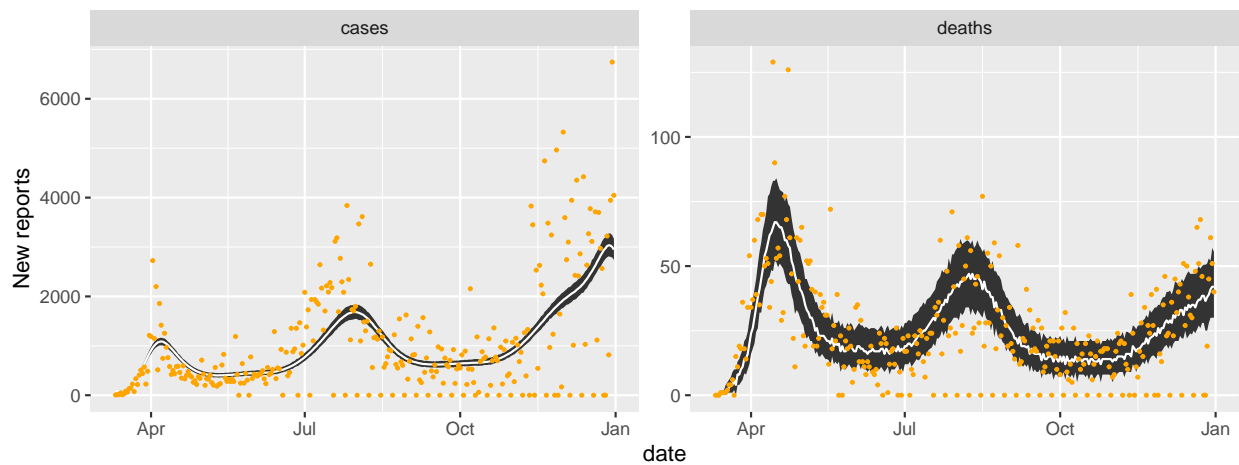
Kansas



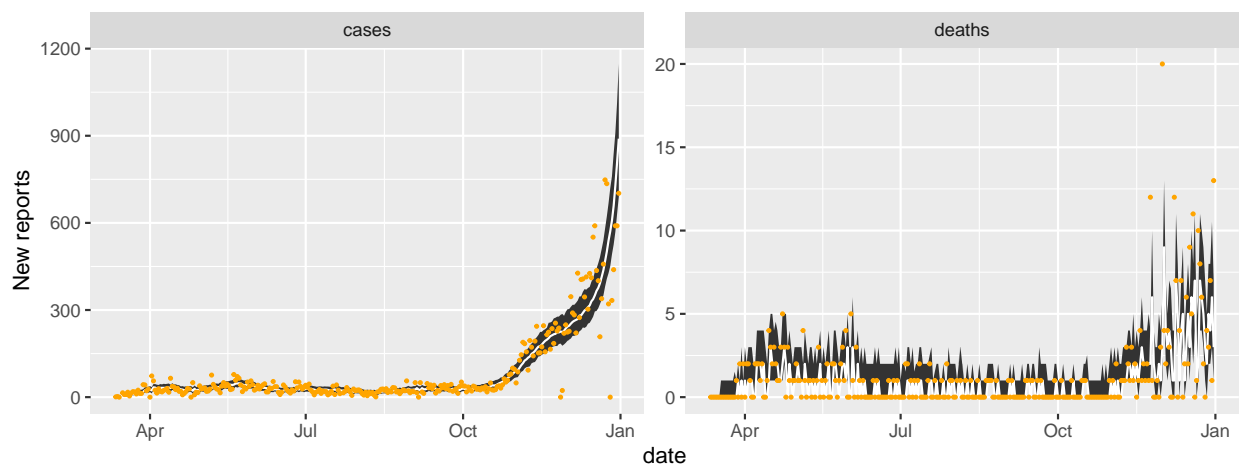
Kentucky



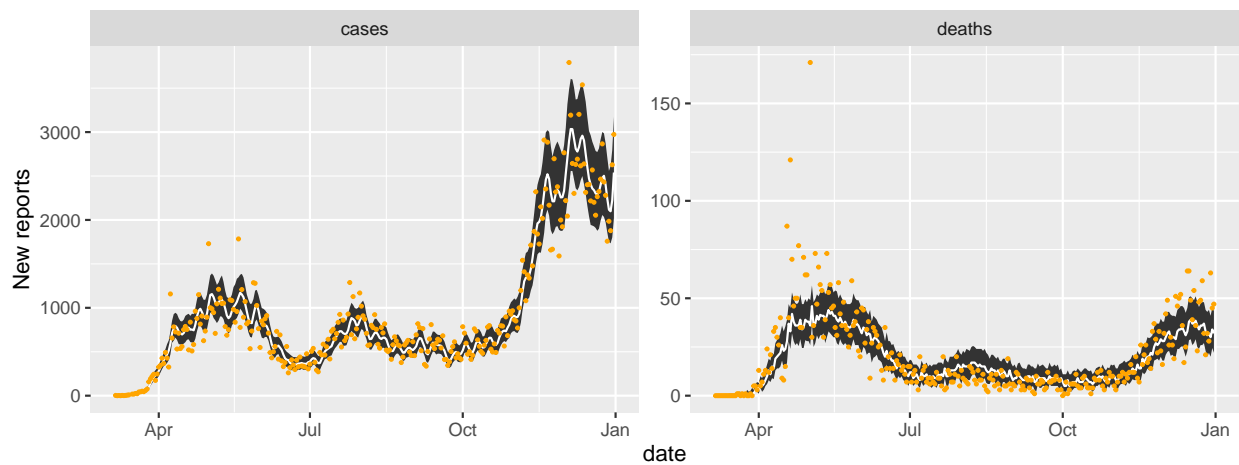
Louisiana



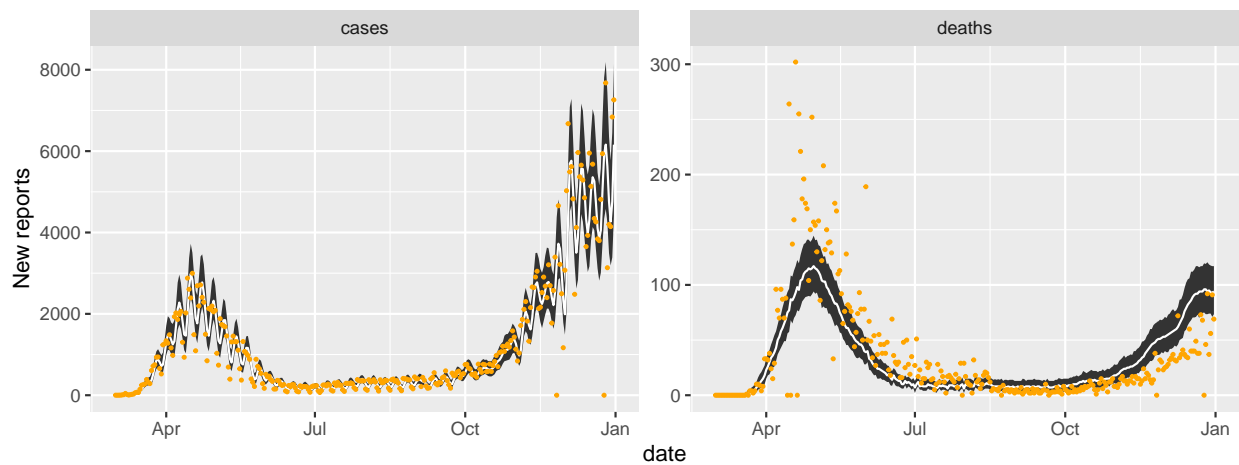
Maine



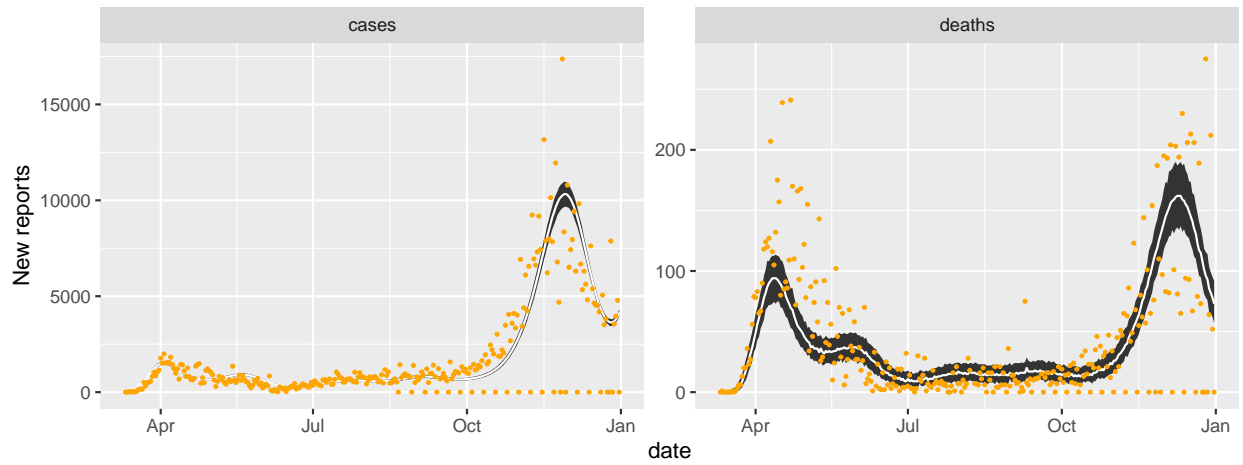
Maryland



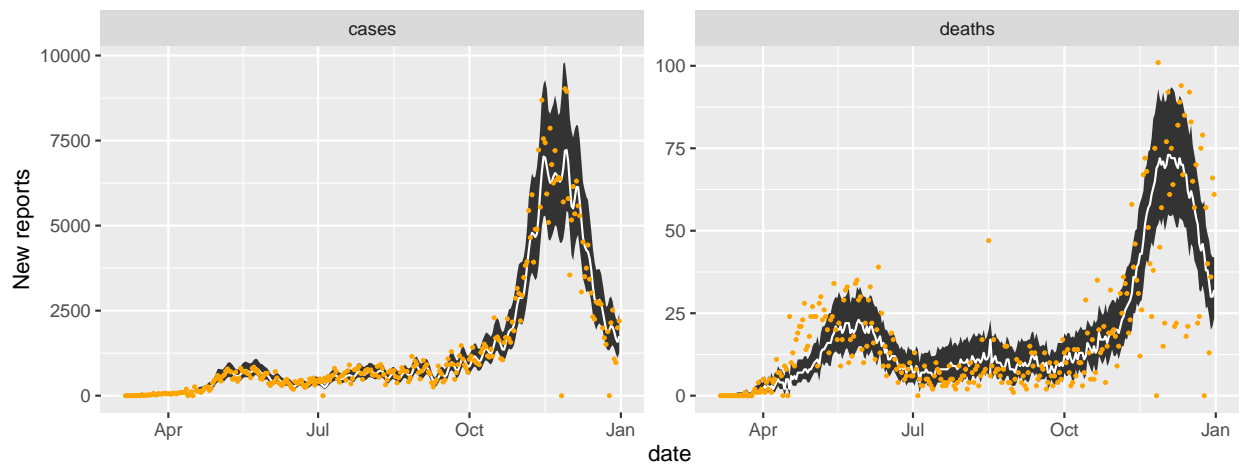
Massachusetts



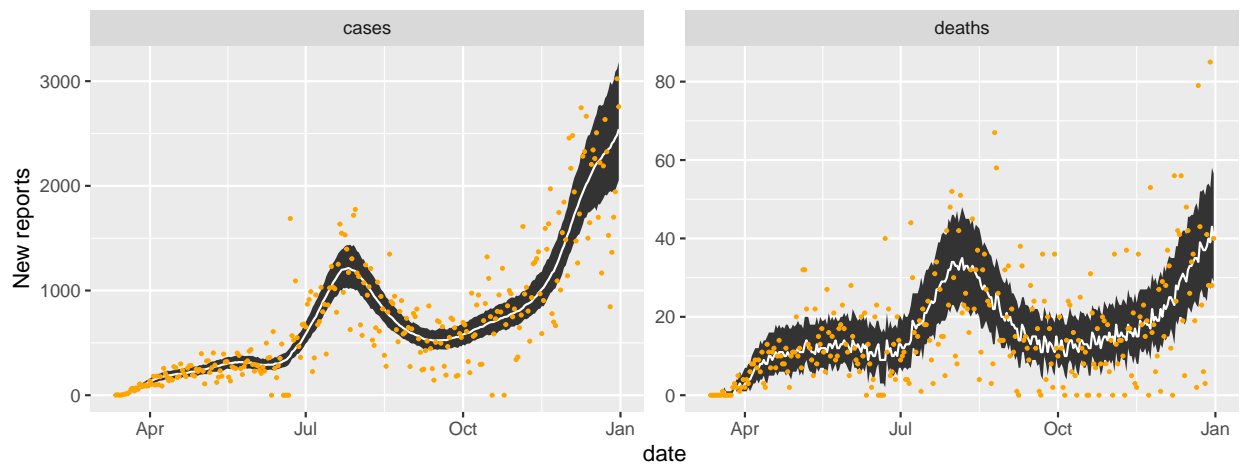
Michigan



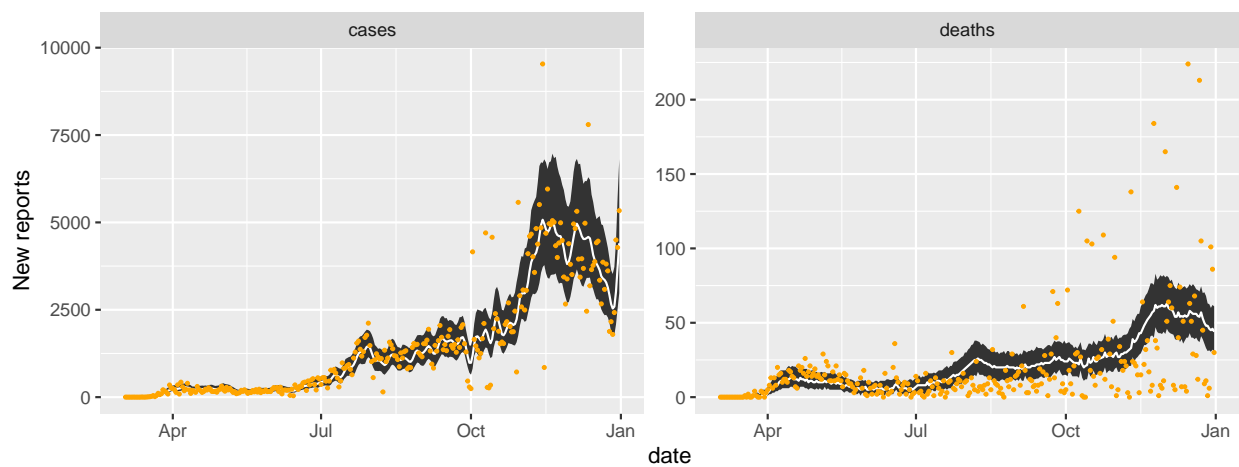
Minnesota



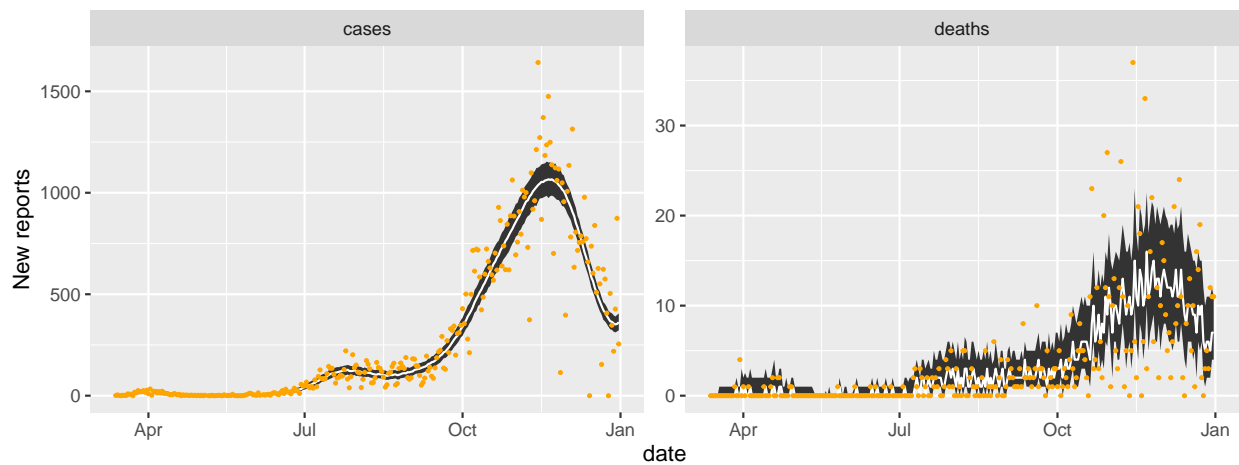
Mississippi



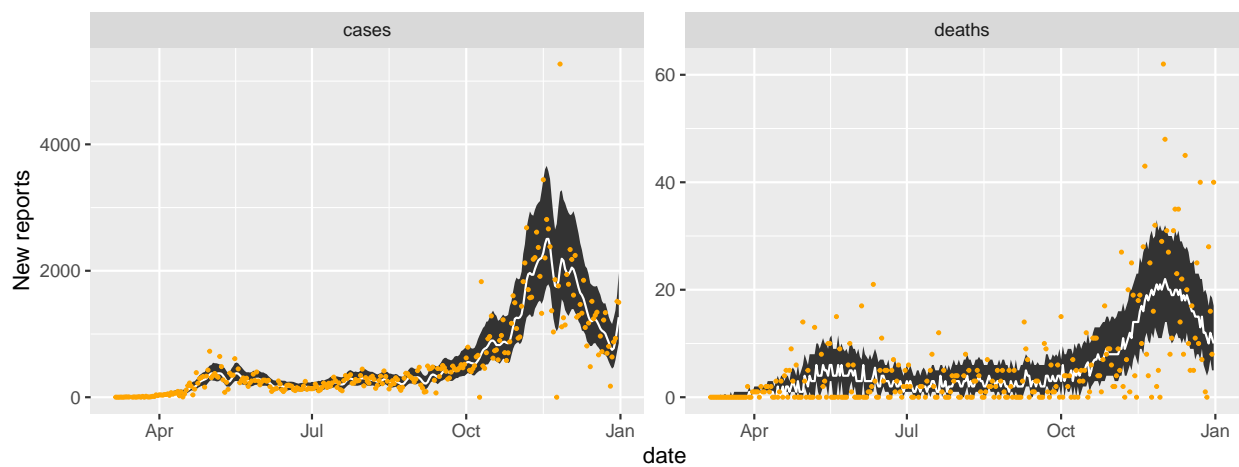
Missouri



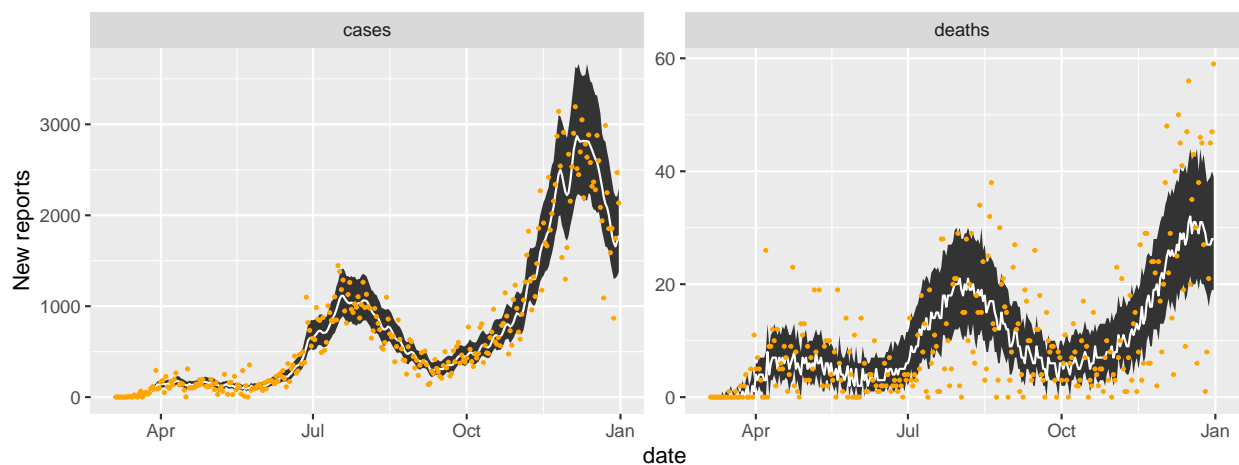
Montana



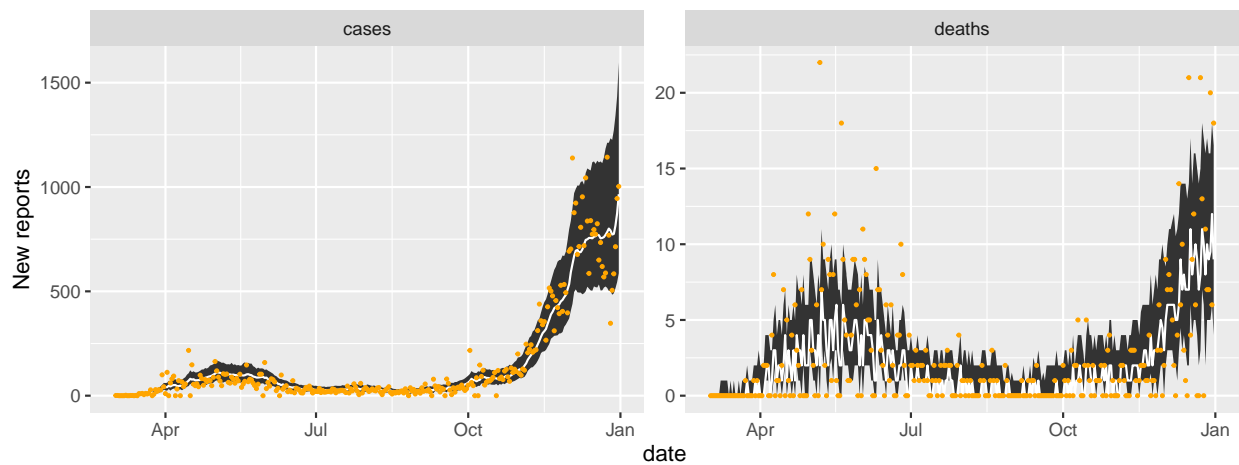
Nebraska



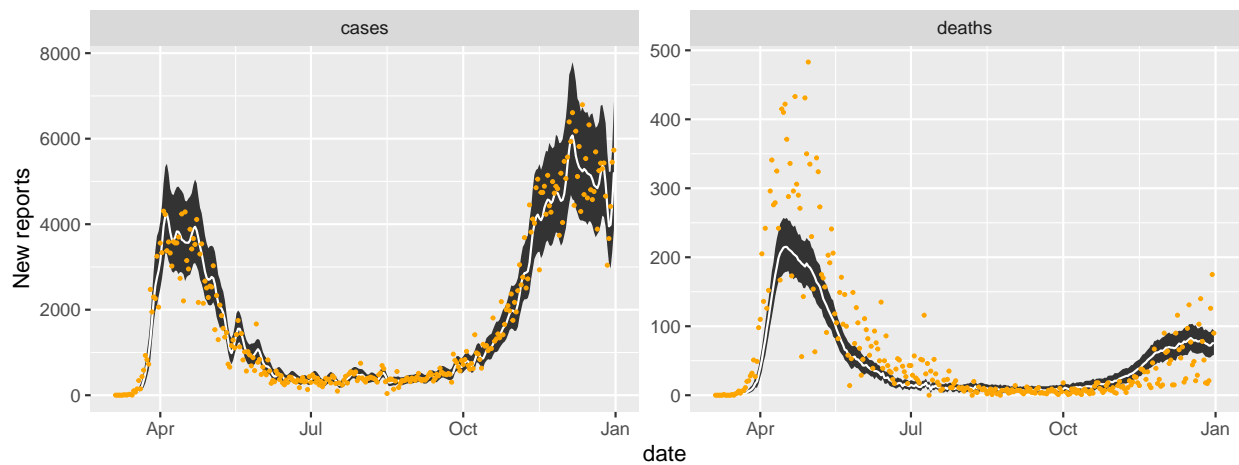
Nevada



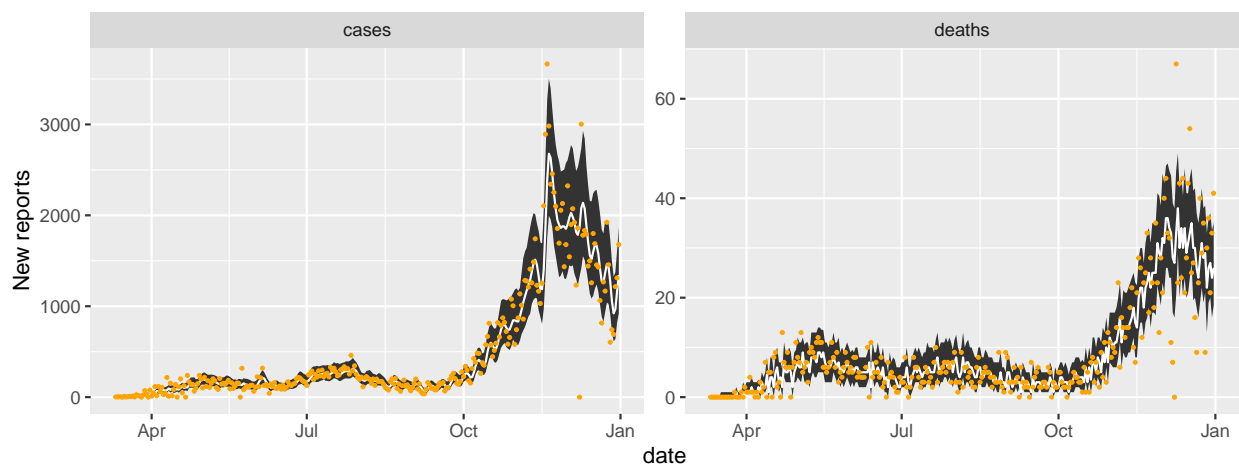
New Hampshire



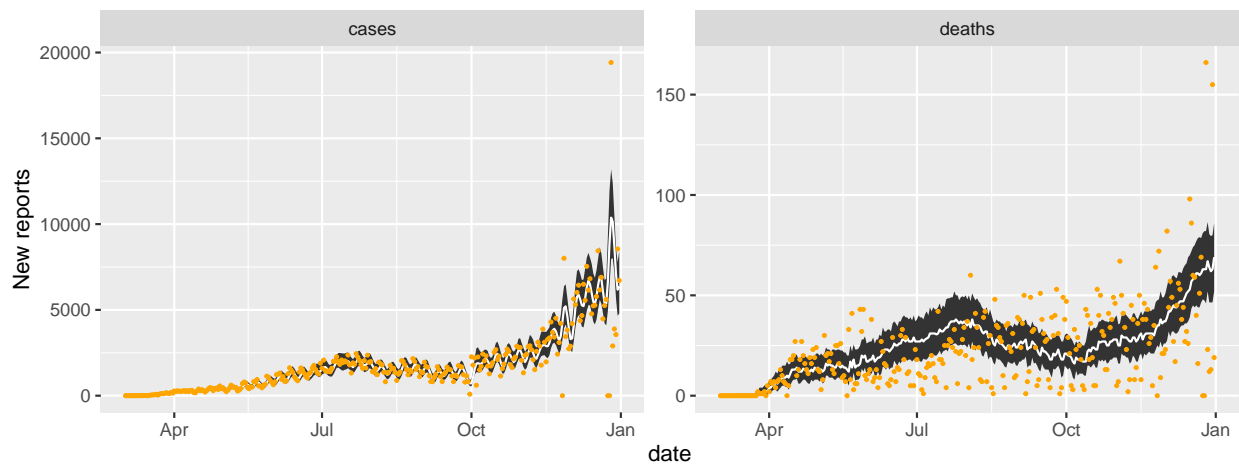
New Jersey



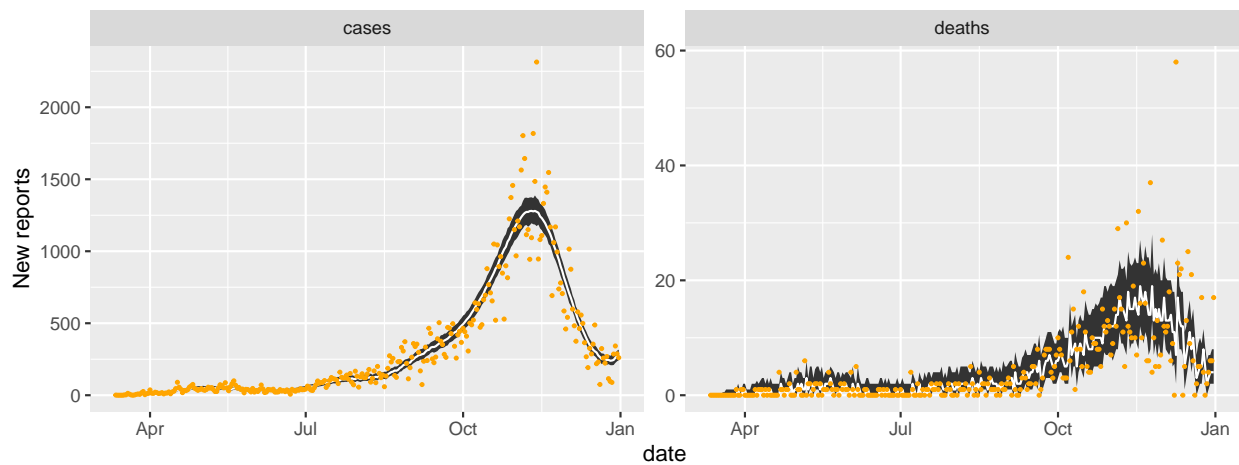
New Mexico



North Carolina

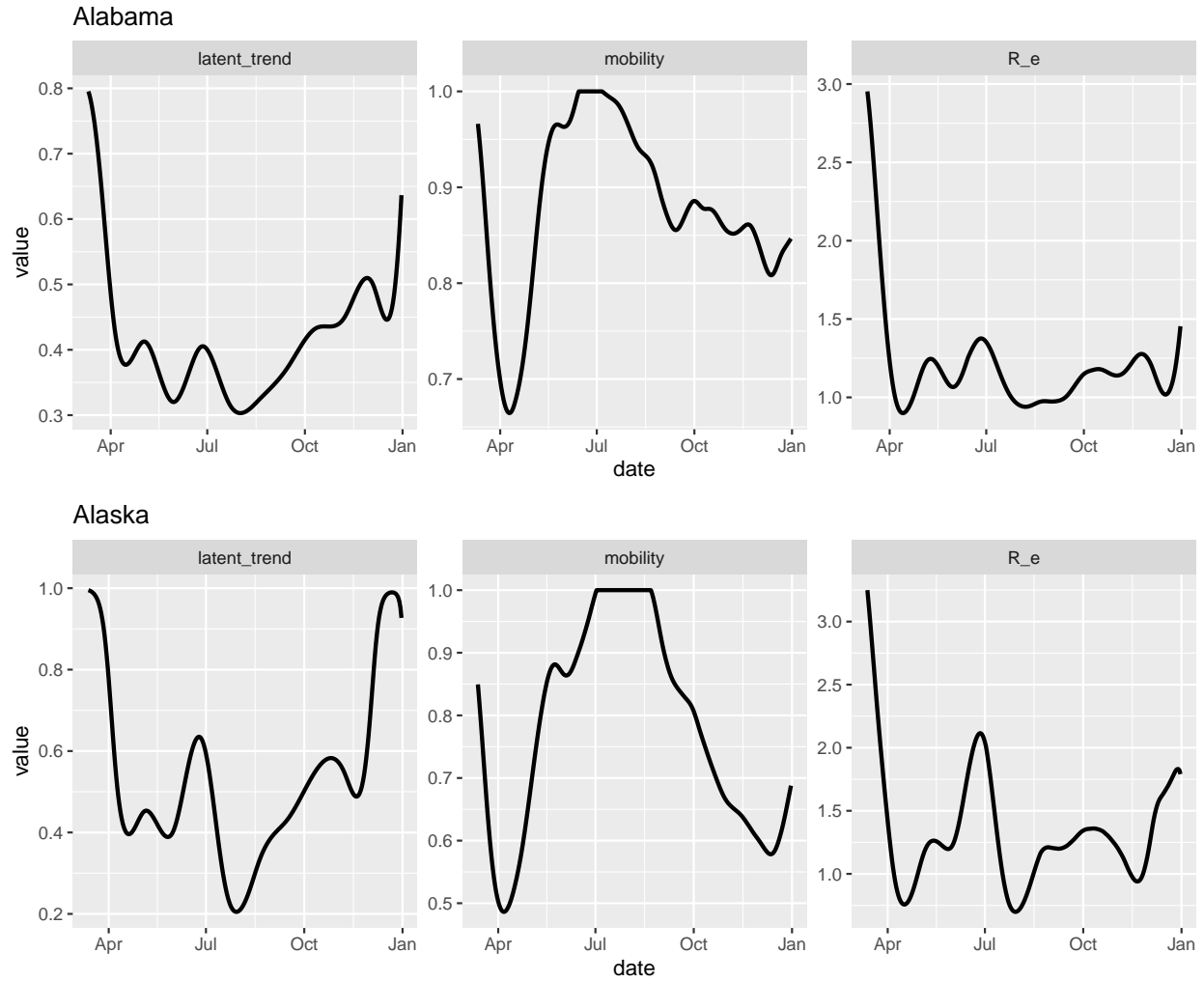


North Dakota

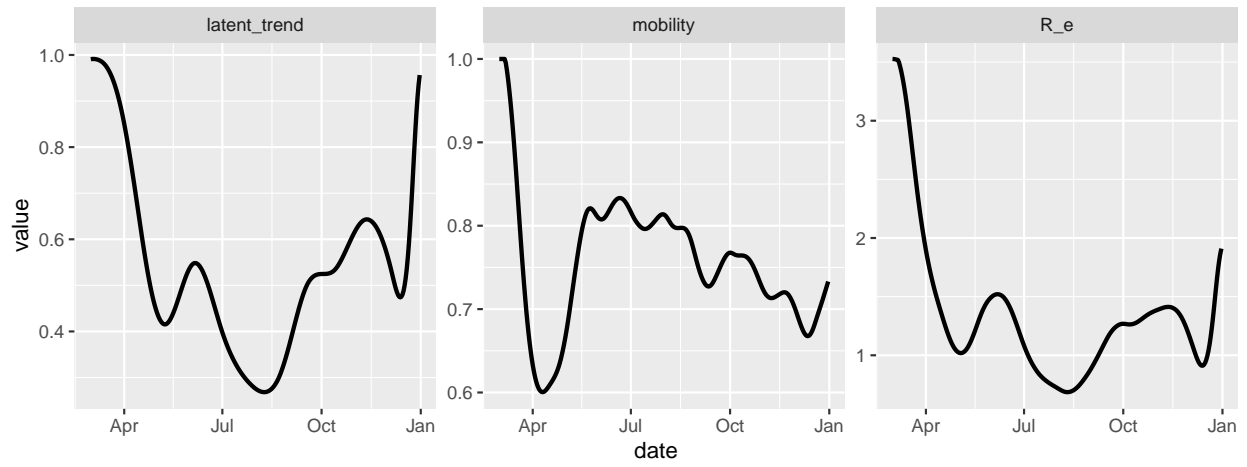


Time series of mobility, estimated latent trend, and \mathcal{R}_e

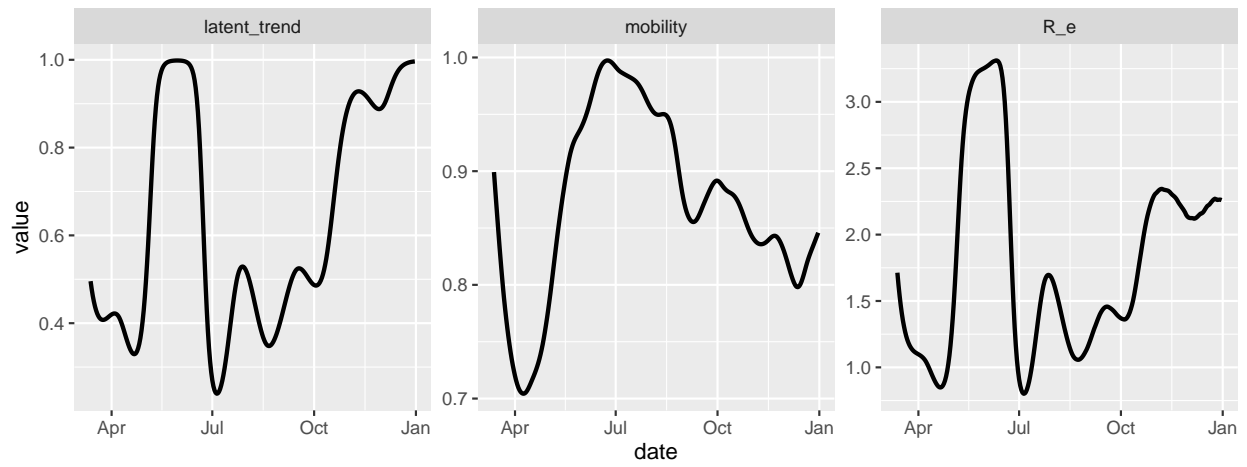
The following plots show the estimated latent trend, relative mobility, and estimated \mathcal{R}_e over time for each state. The latent trend is estimated using the maximum likelihood parameter estimates for the g_i coefficients of the B-spline (see Materials and Methods in main text).



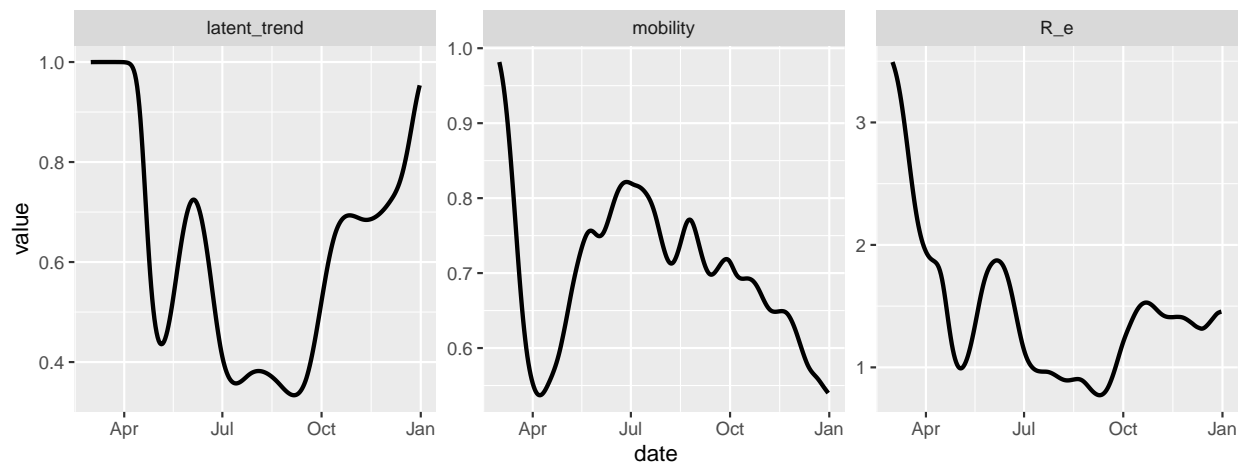
Arizona



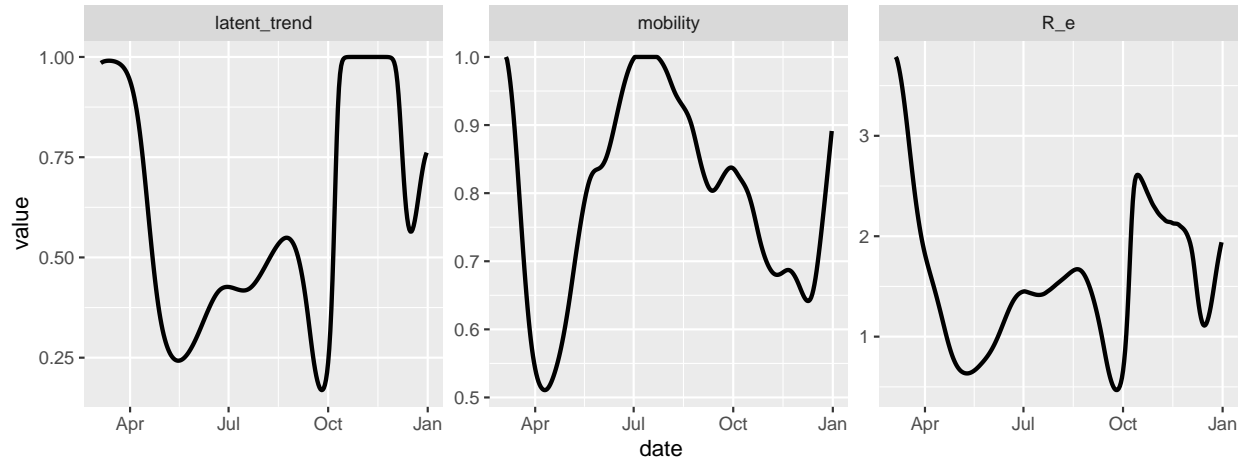
Arkansas



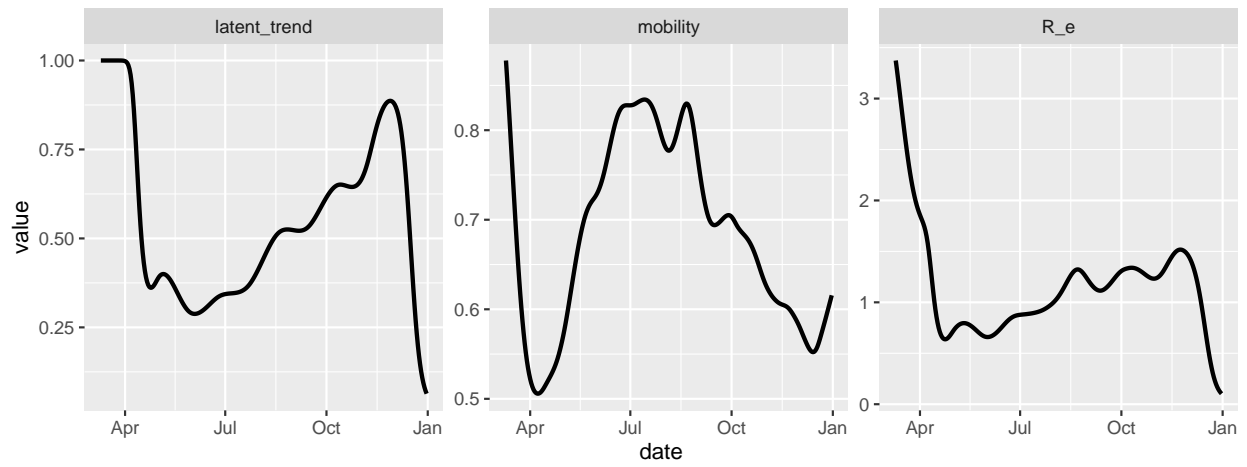
California



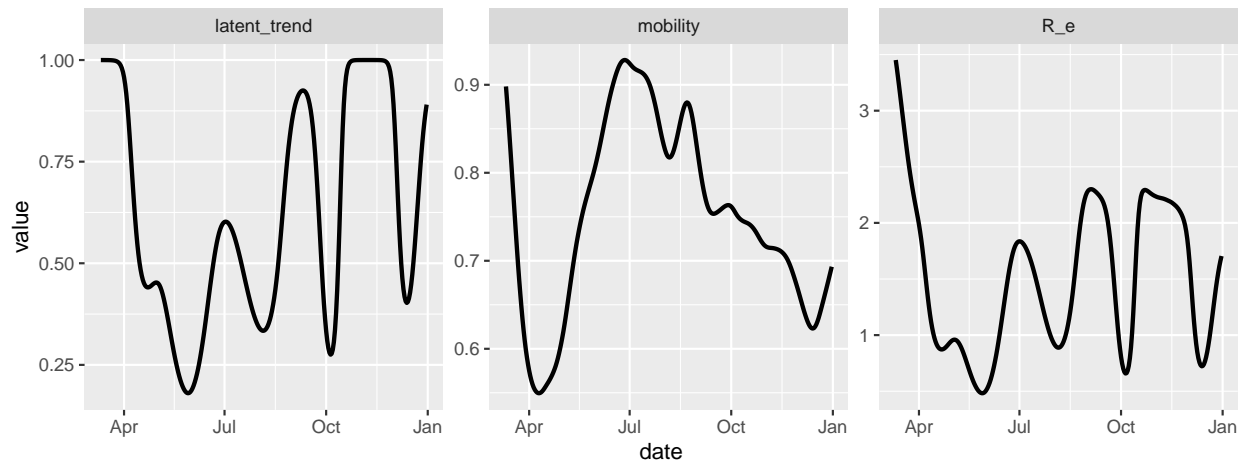
Colorado



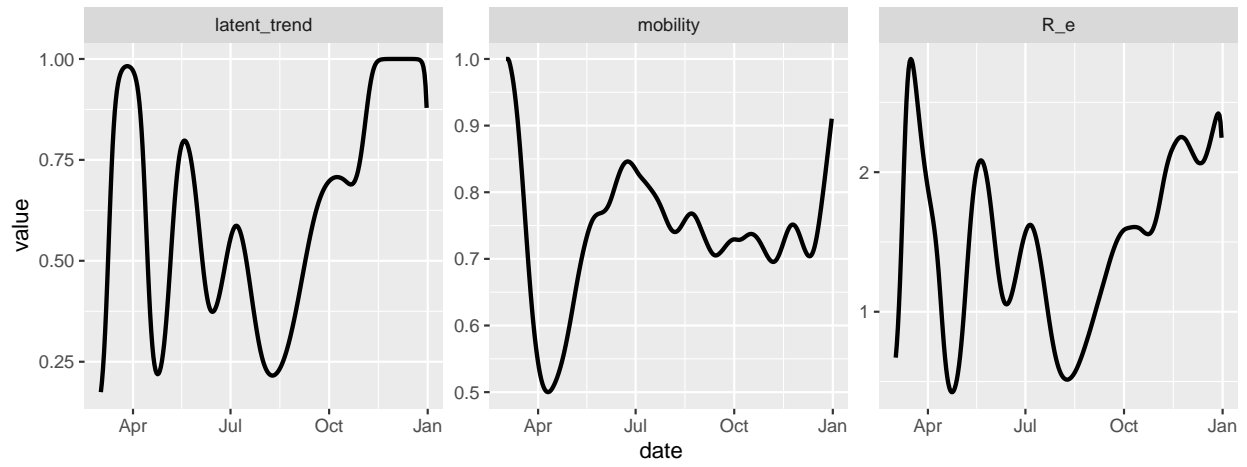
Connecticut



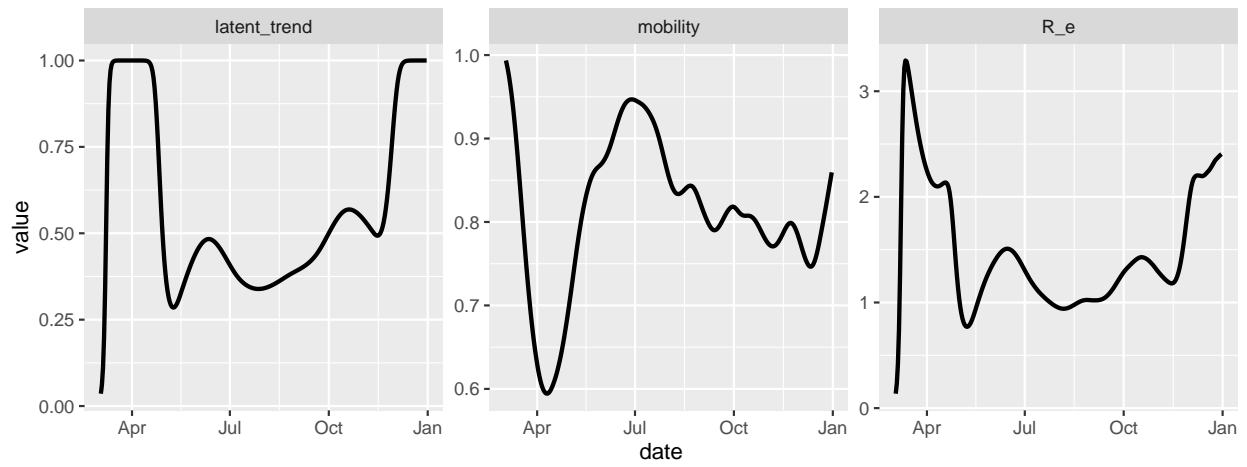
Delaware



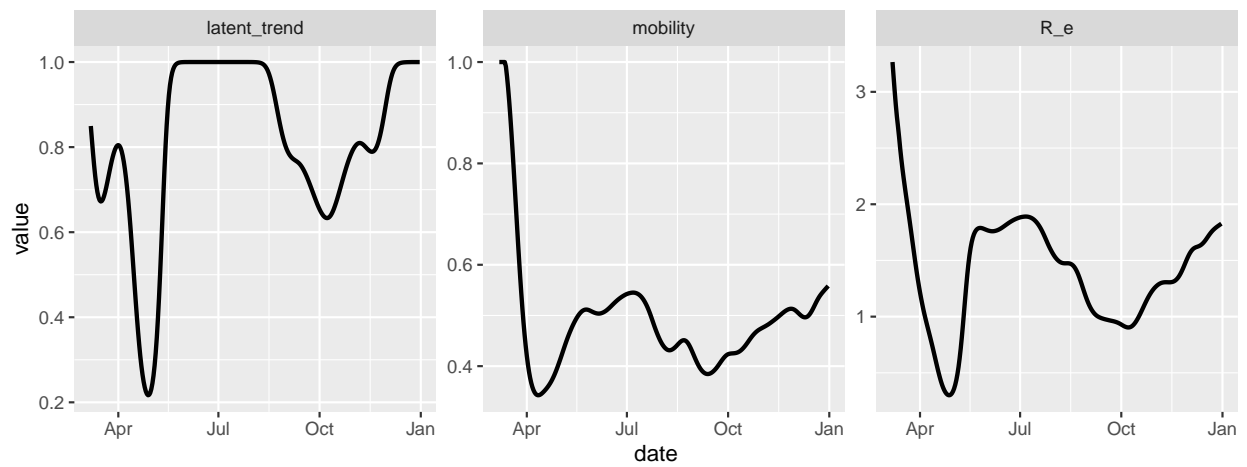
Florida



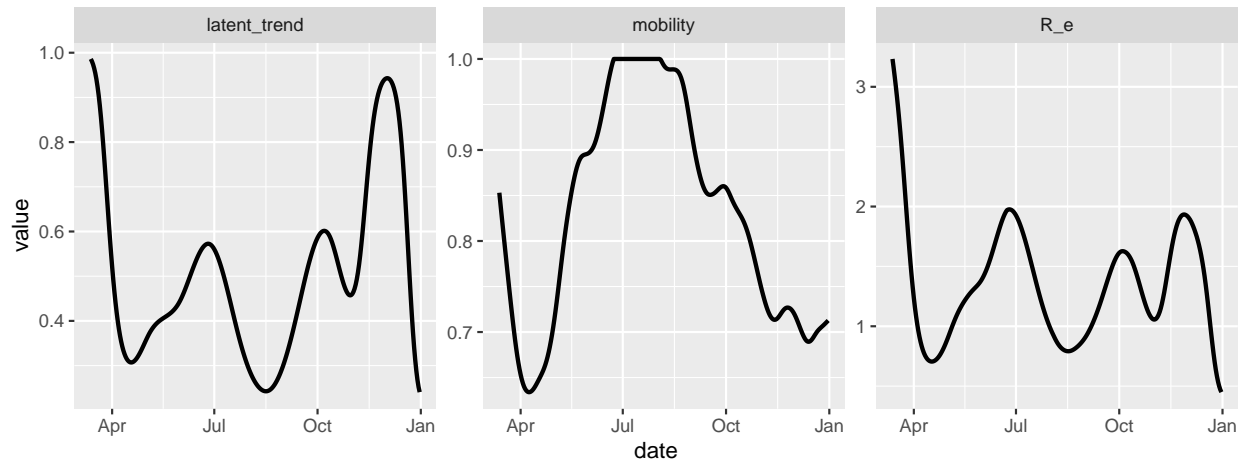
Georgia



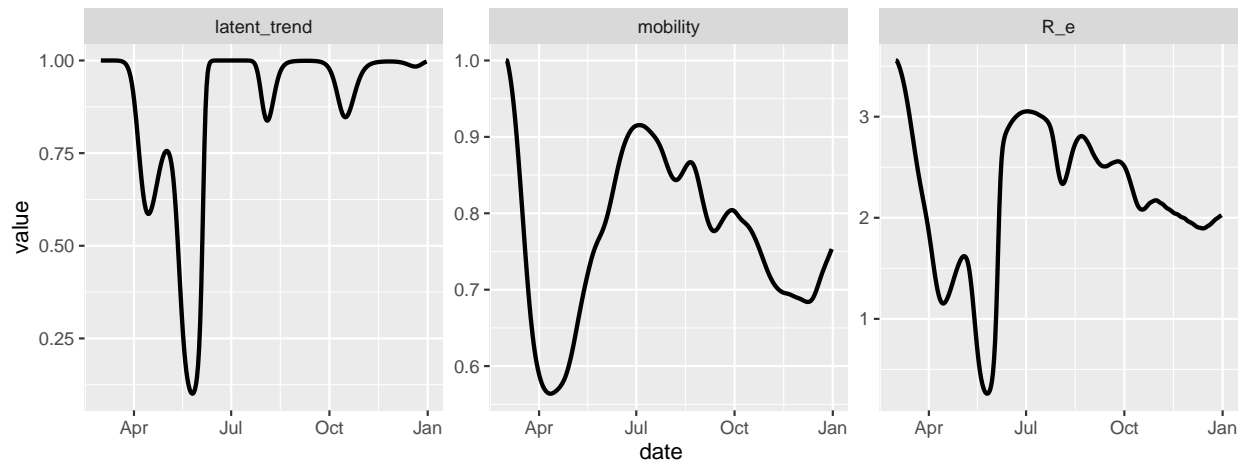
Hawaii



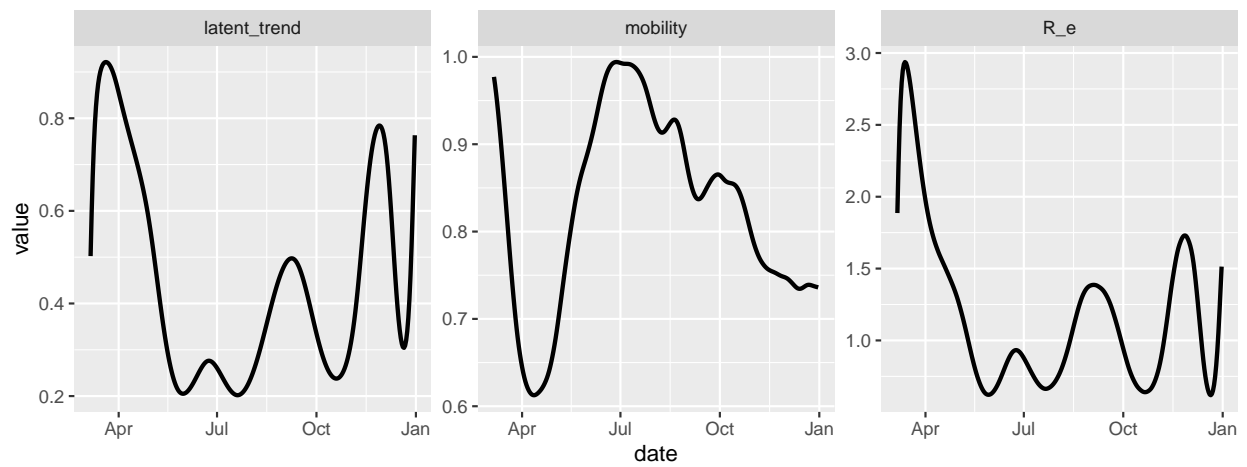
Idaho

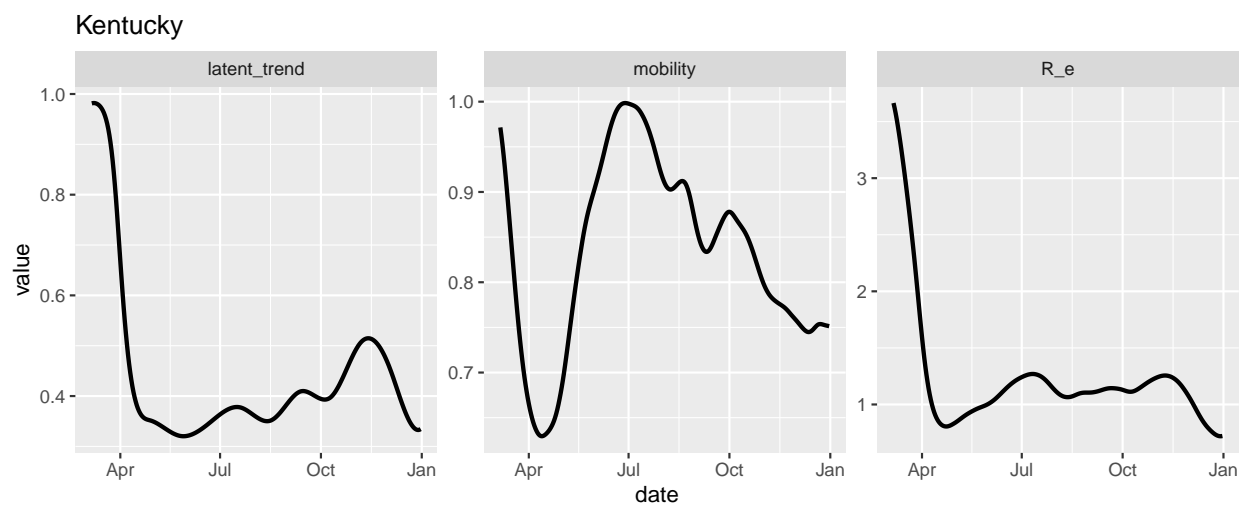
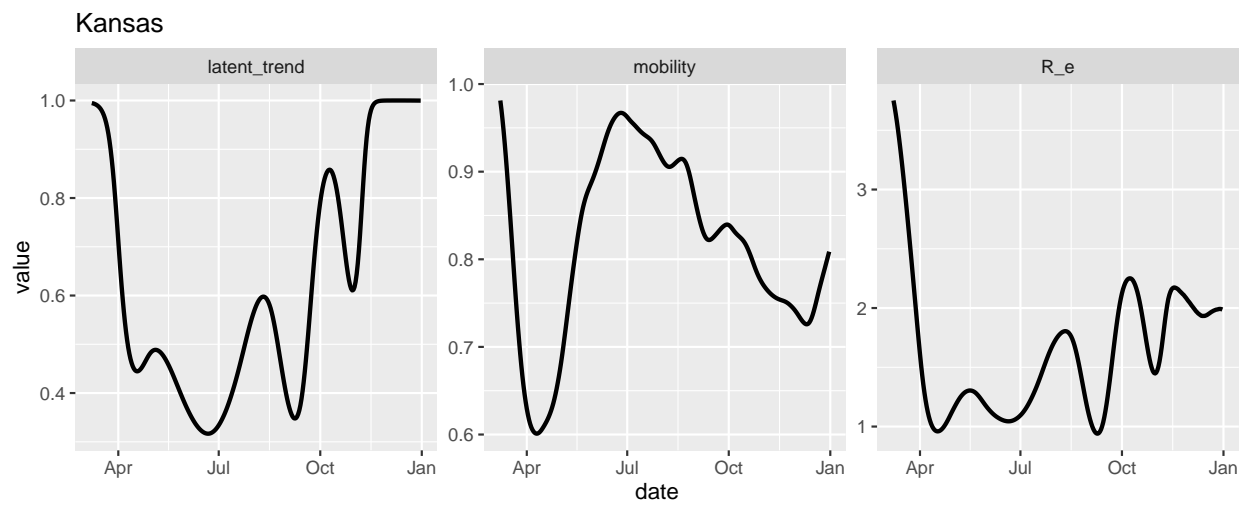
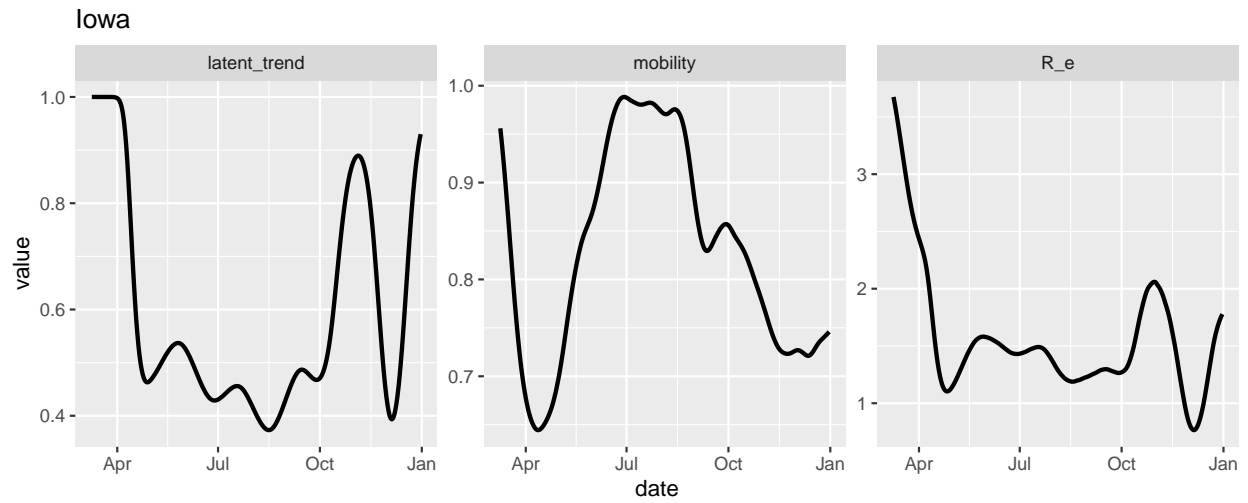


Illinois

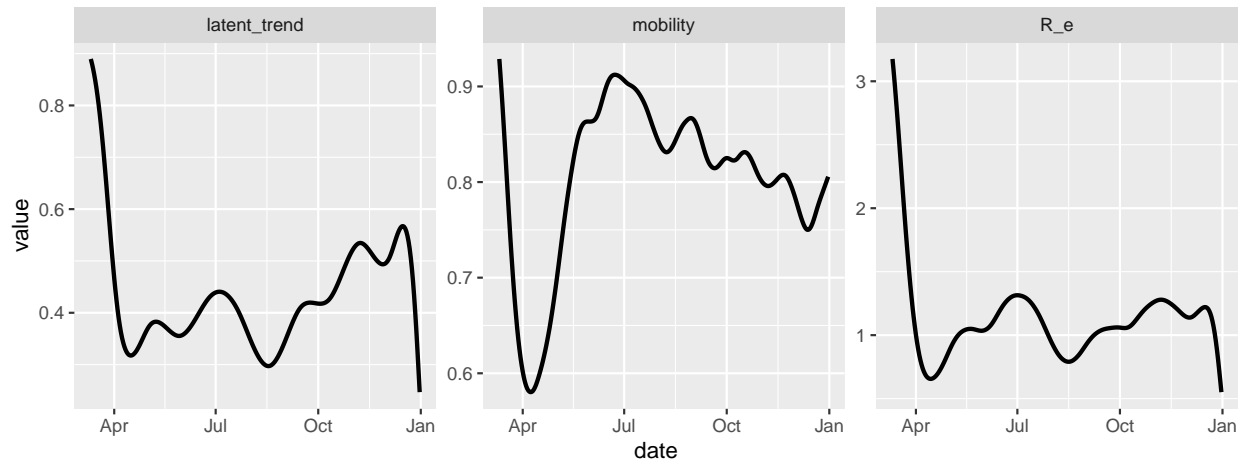


Indiana

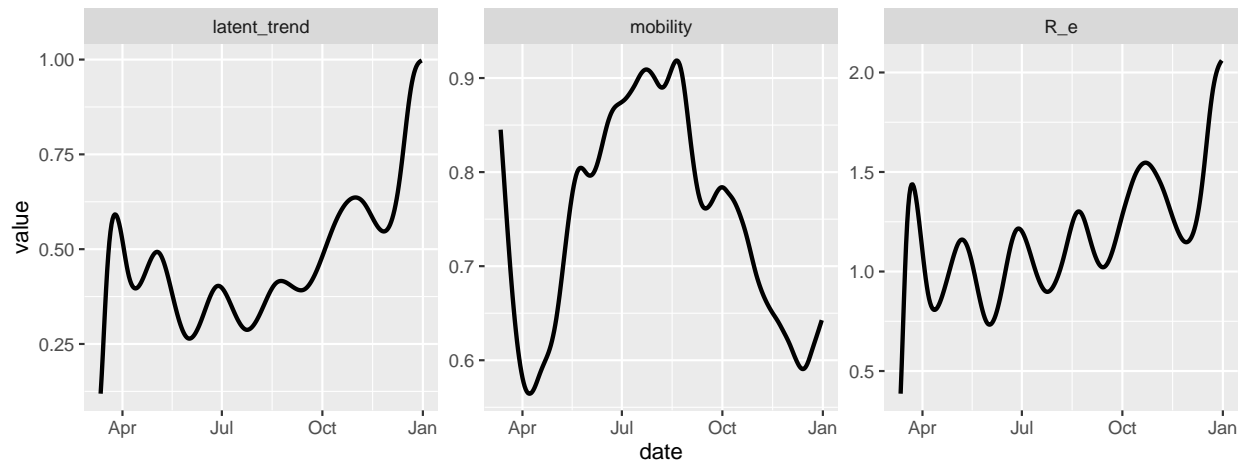




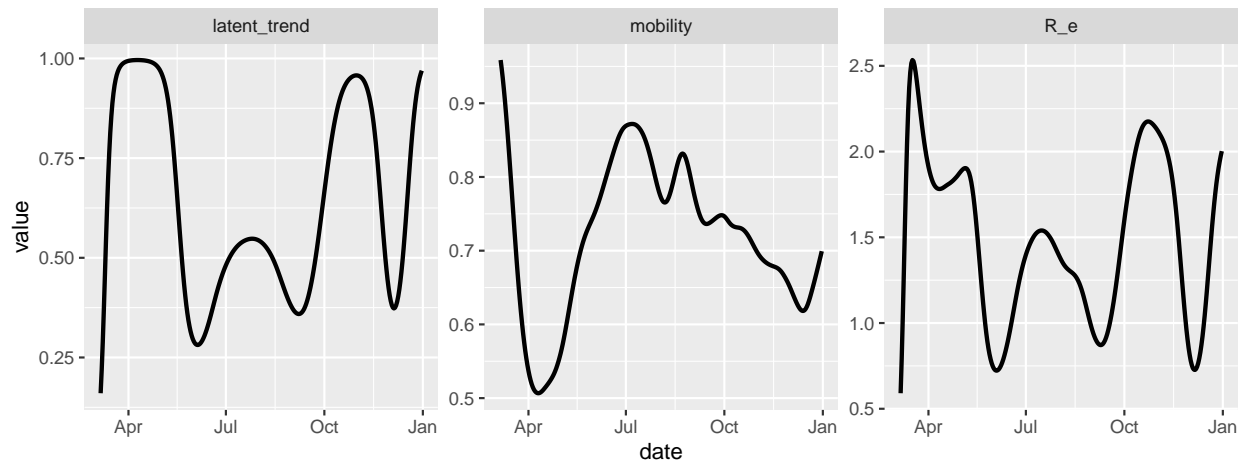
Louisiana



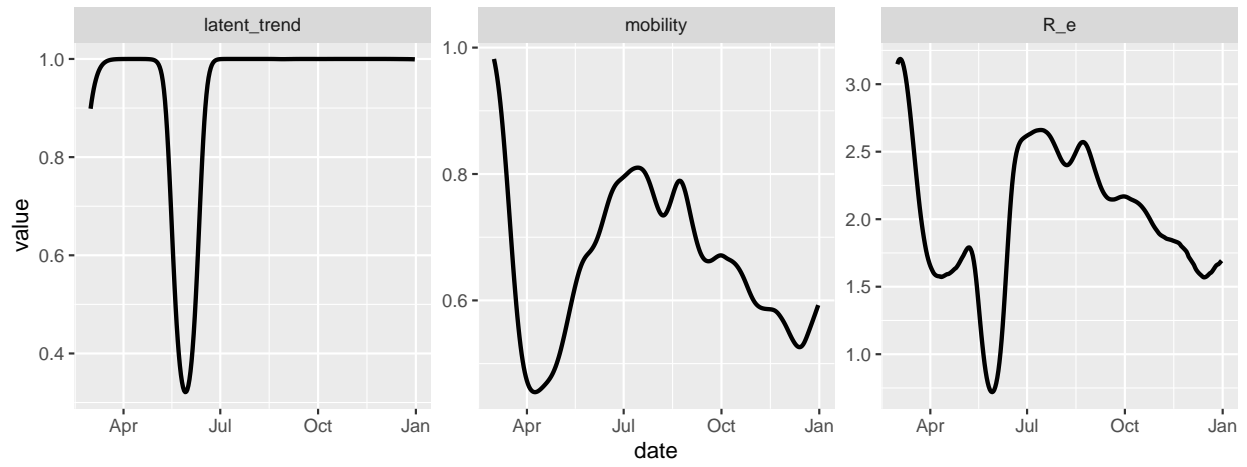
Maine



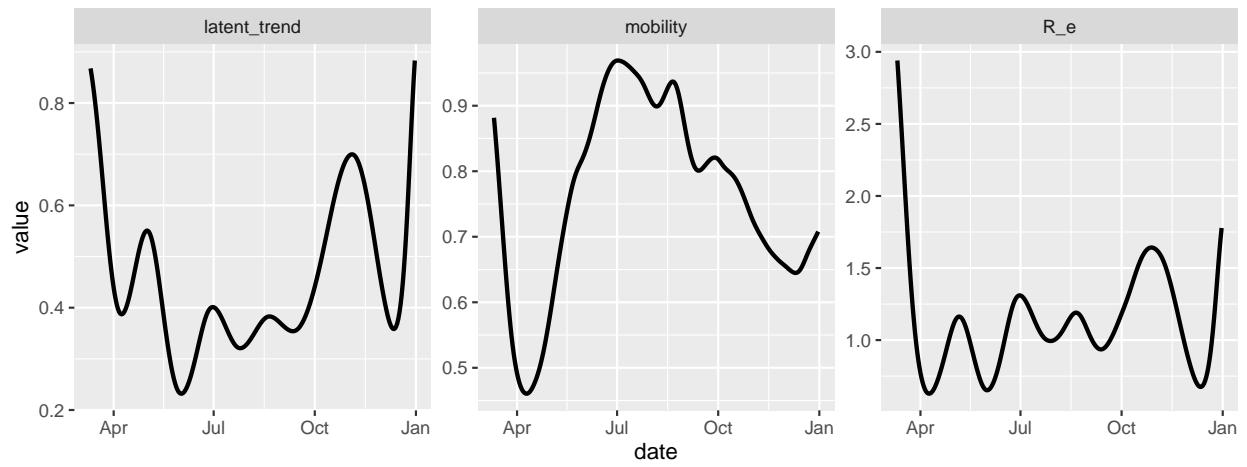
Maryland



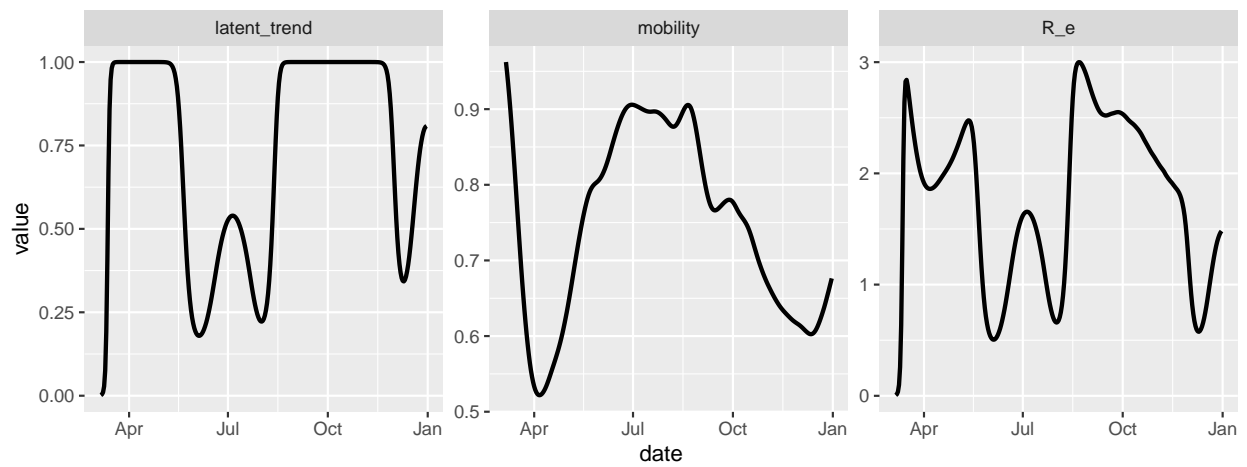
Massachusetts



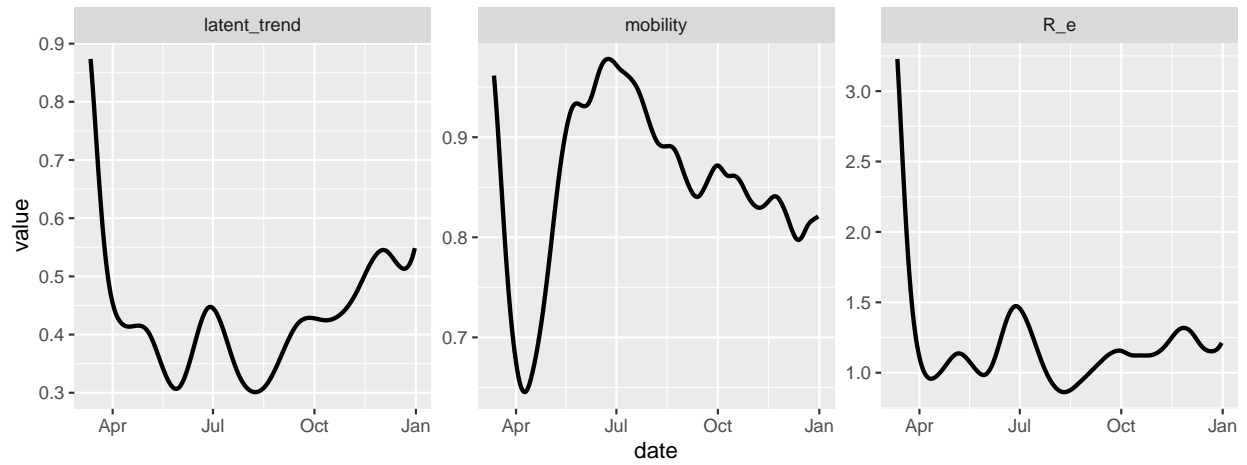
Michigan



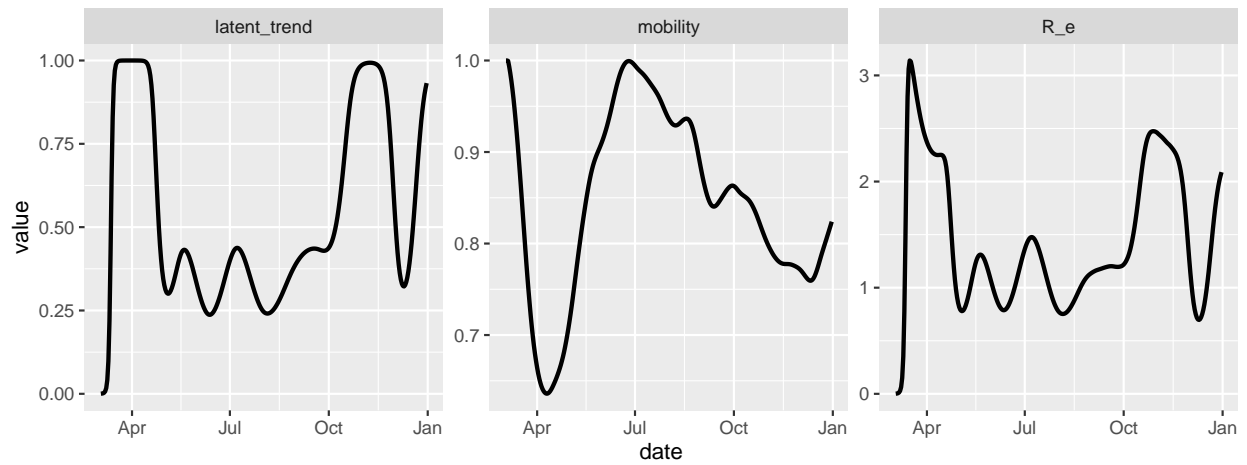
Minnesota



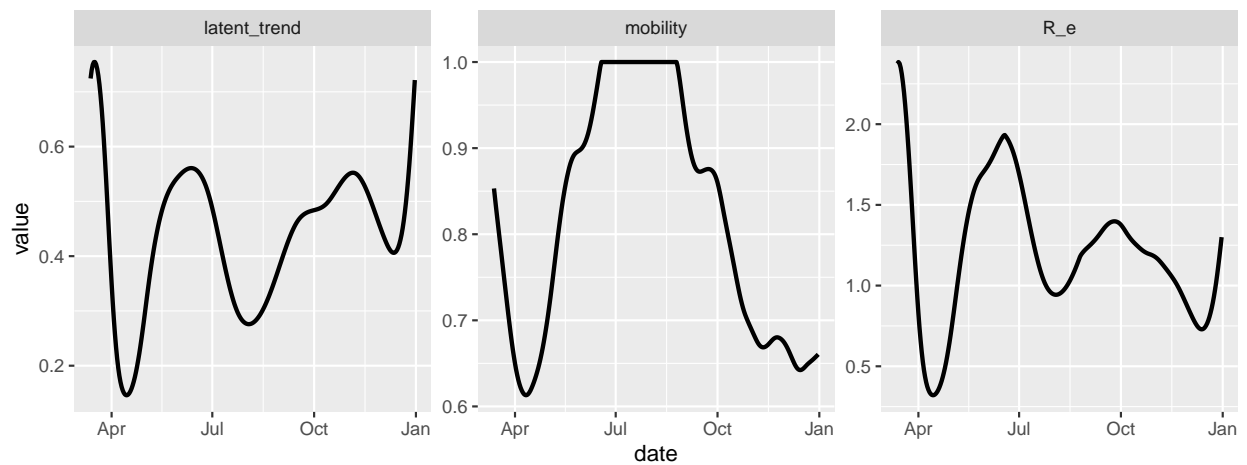
Mississippi



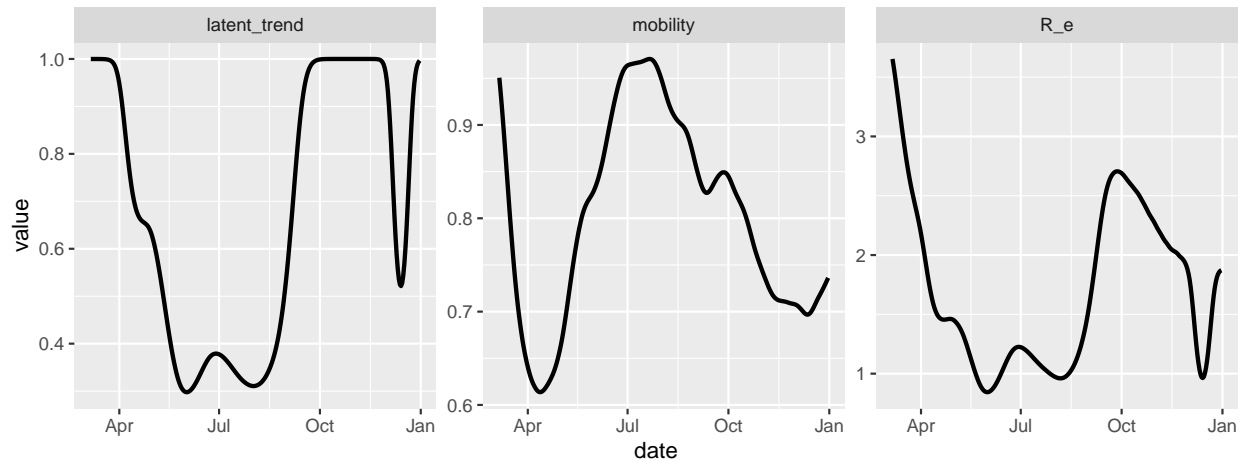
Missouri



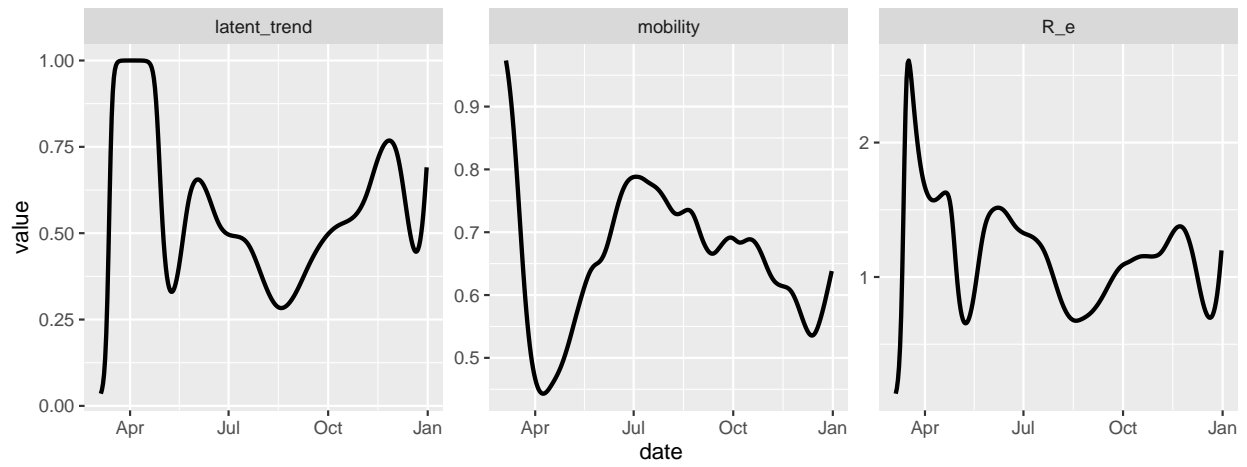
Montana



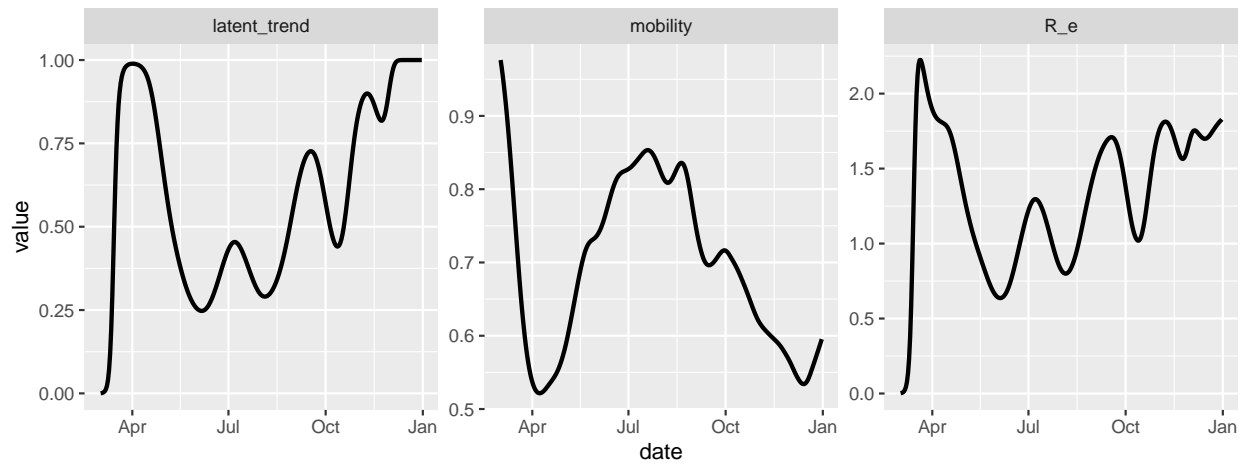
Nebraska



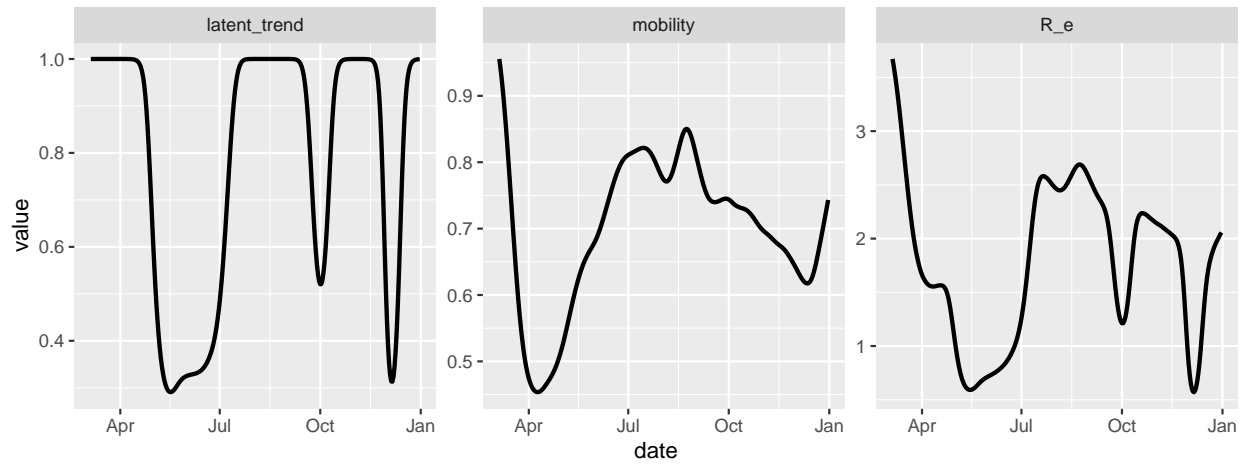
Nevada



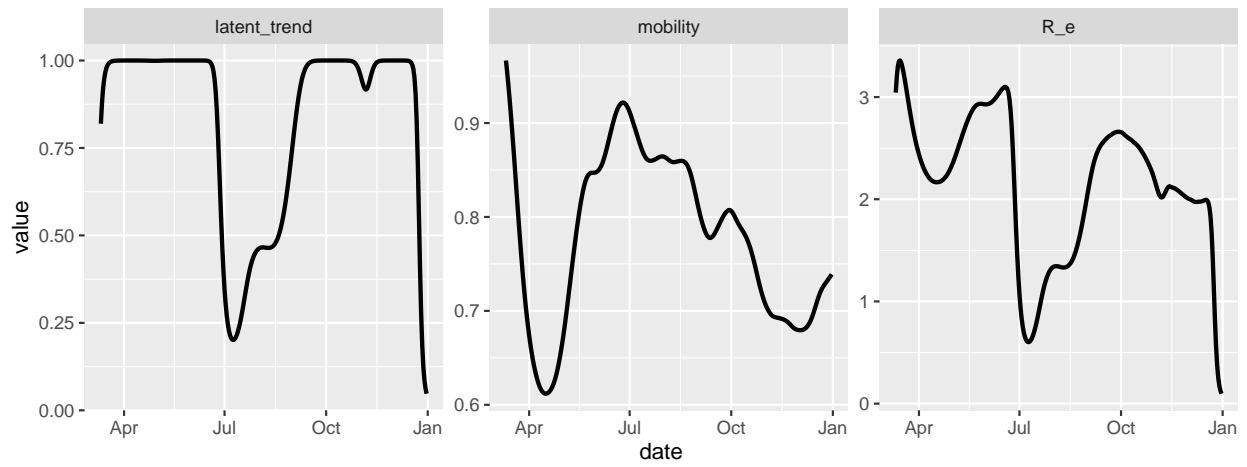
New Hampshire



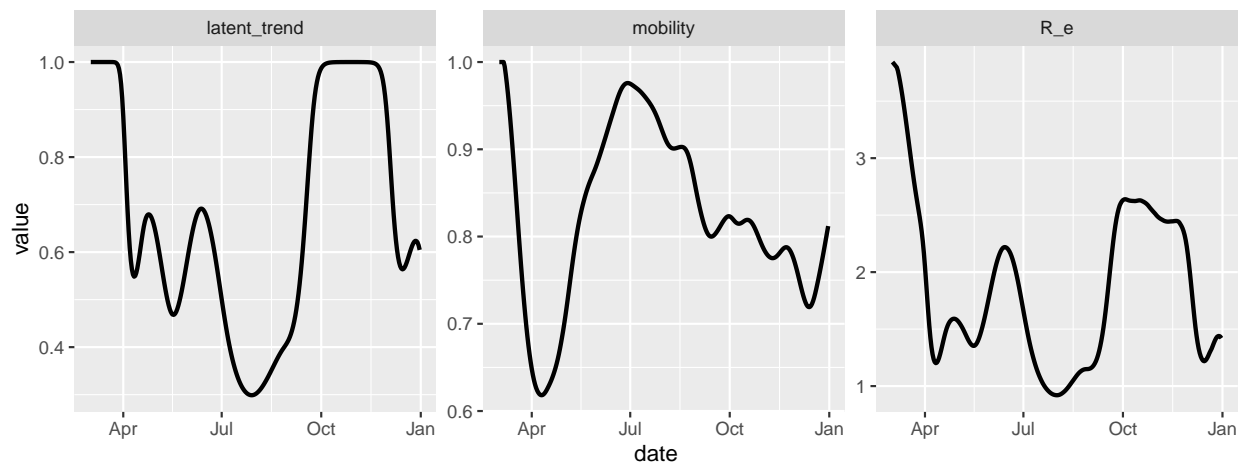
New Jersey



New Mexico



North Carolina



North Dakota

