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

2022-08-02

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

Model diagram

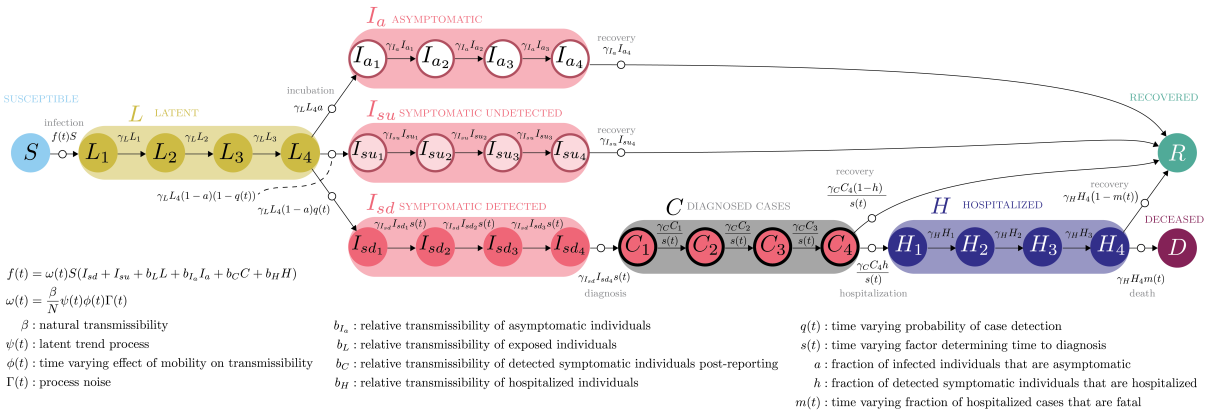


Figure S1: Model diagram.

Model parameters

Table S1: Fixed model parameters.

Parameter definition	Parameter symbol	Value	Source
Reproduction number	$\mathcal{R}_0$	7	Sanche et al. (2020)
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)

Parameter definition	Parameter symbol	Value	Source
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_{I_a}	0.27	Assumption
Relative transmissibility of H to $I_{su/sd}$	b_{I_a}	4.5×10^{-5}	Moghadas et al. (2020)
Rate of movement through a single L stage	γ_L	$\frac{1}{1 \text{ days}}$	Li et al. (2020)
Rate of movement through a single I_a stage	γ_{I_a}	$\frac{1}{0.88 \text{ days}}$	Li et al. (2020)
Rate of movement through a single I_{su} stage	$\gamma_{I_{su}}$	$\frac{1}{1.5 \text{ days}}$	Li et al. (2020)
Rate of movement through a single I_{sd} stage	$\gamma_{I_{su}}$	$\frac{1}{0.75 \text{ days}}$	Assumption
Rate of movement through a single C stage	γ_C	$\frac{1}{0.75 \text{ days}}$	Assumption
Rate of movement through a single H stage	γ_H	$\frac{1}{1.5 \text{ days}}$	Moghadas 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 incrases	$s_{\max}$	1.0	Assumption
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.

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-demogaphic 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]

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.

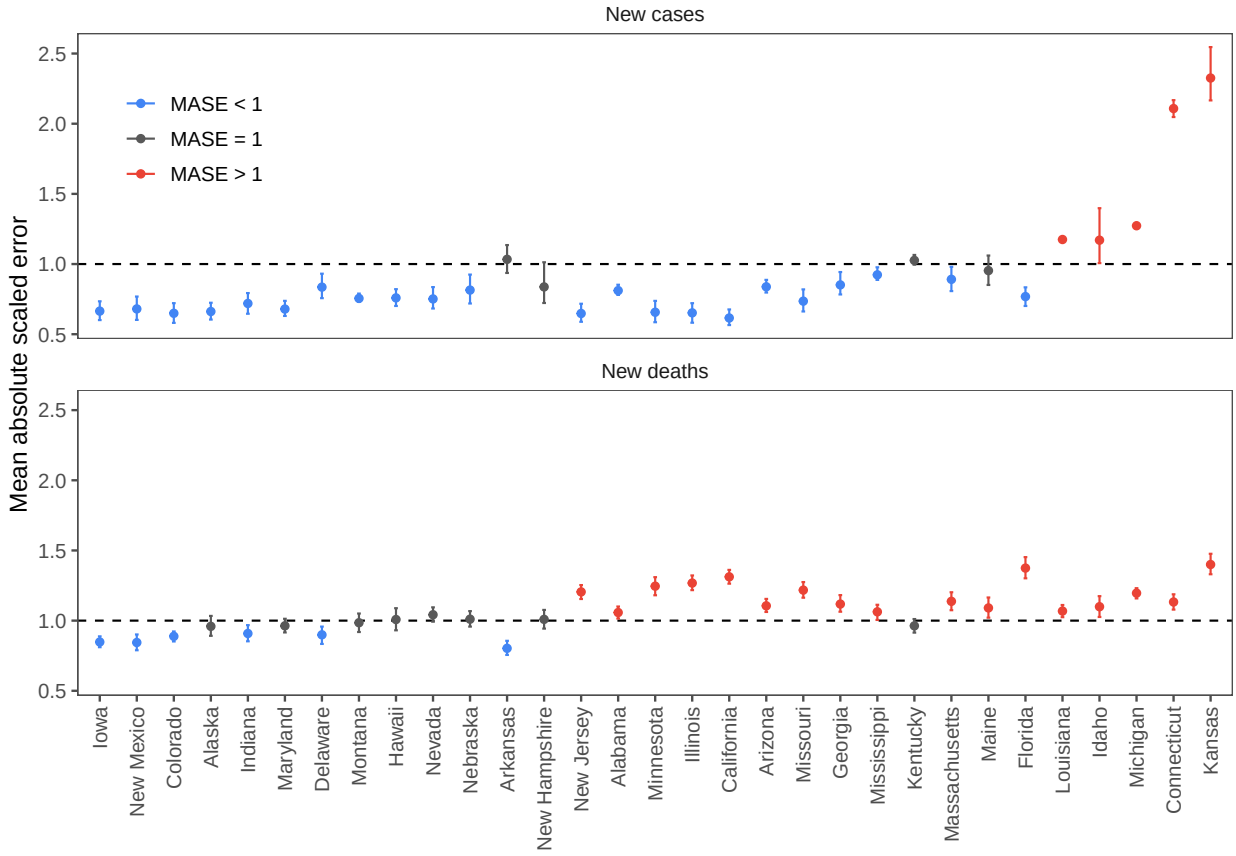
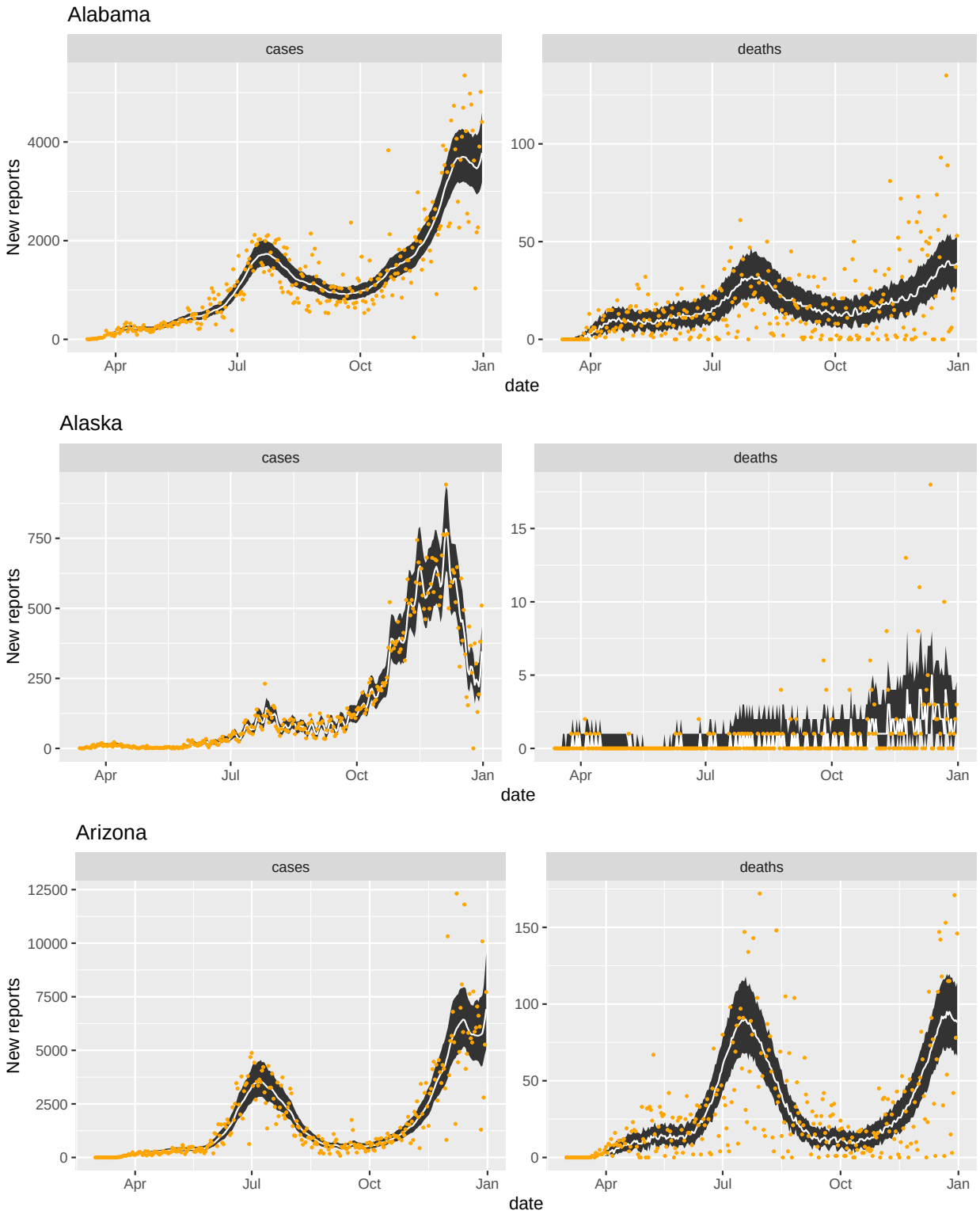
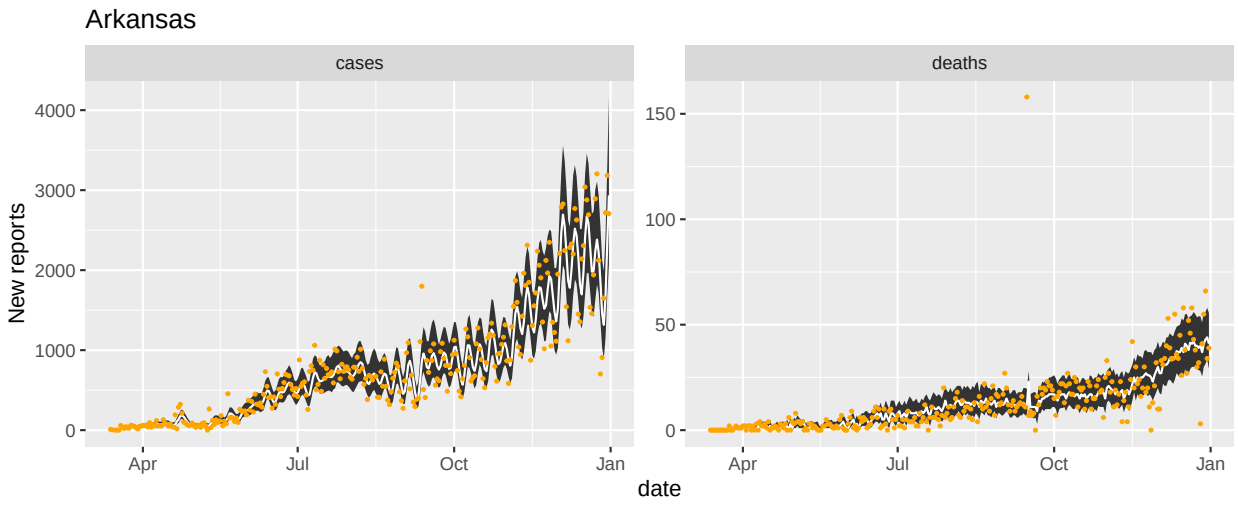


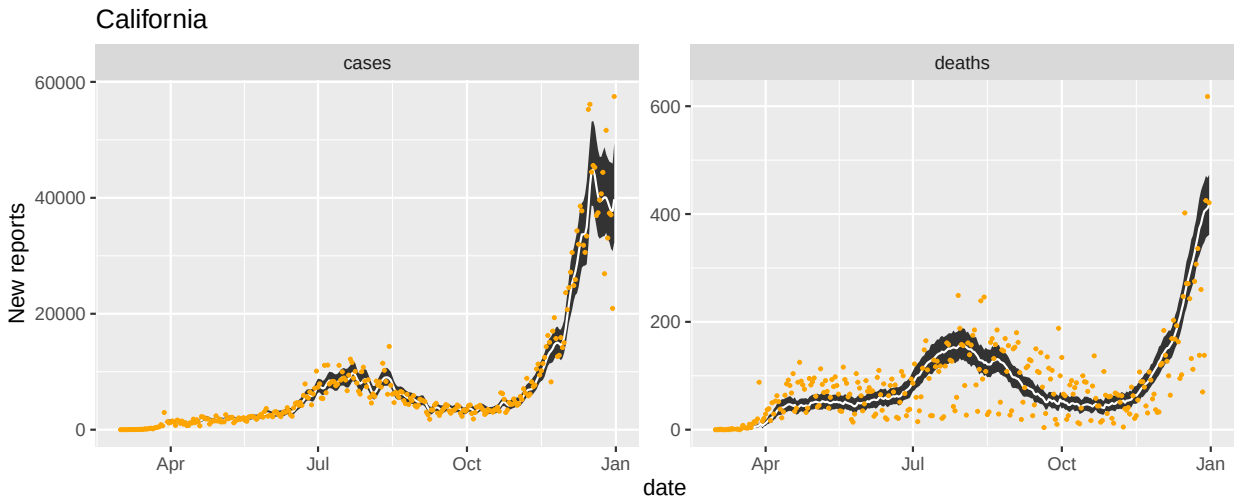
Figure S2: Mean and 95% intervals of mean absolute scaled errors (MASE) for each state. MASE was calculated for each of the 500 particle filter replicates for each state, yielding a distribution of MASE values. The point shows the mean of the MASE distribution and the errorbars show the 95% interval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

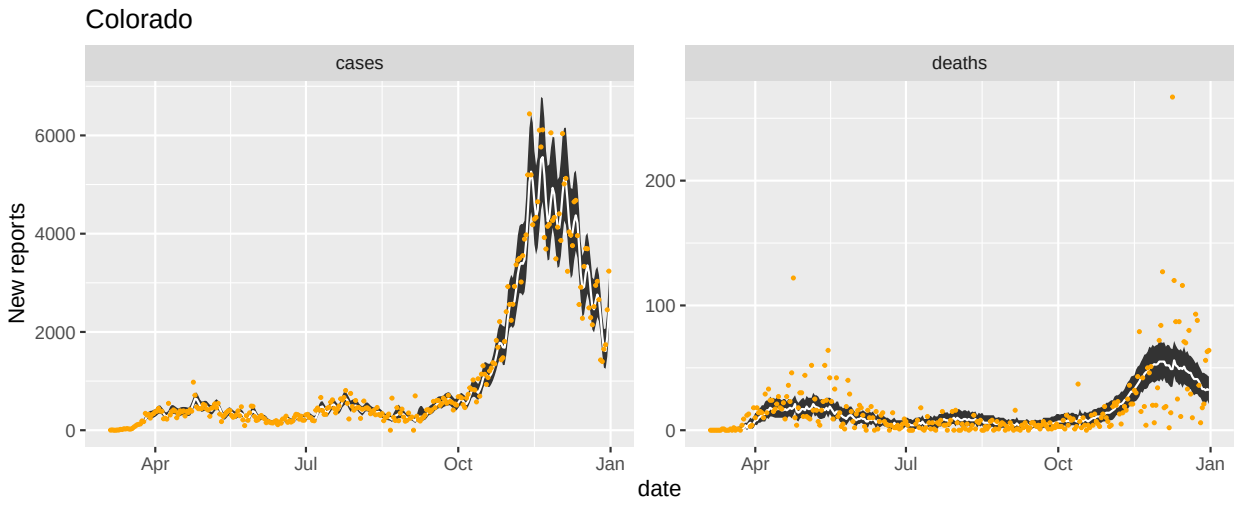
Time series of incident case and death reports

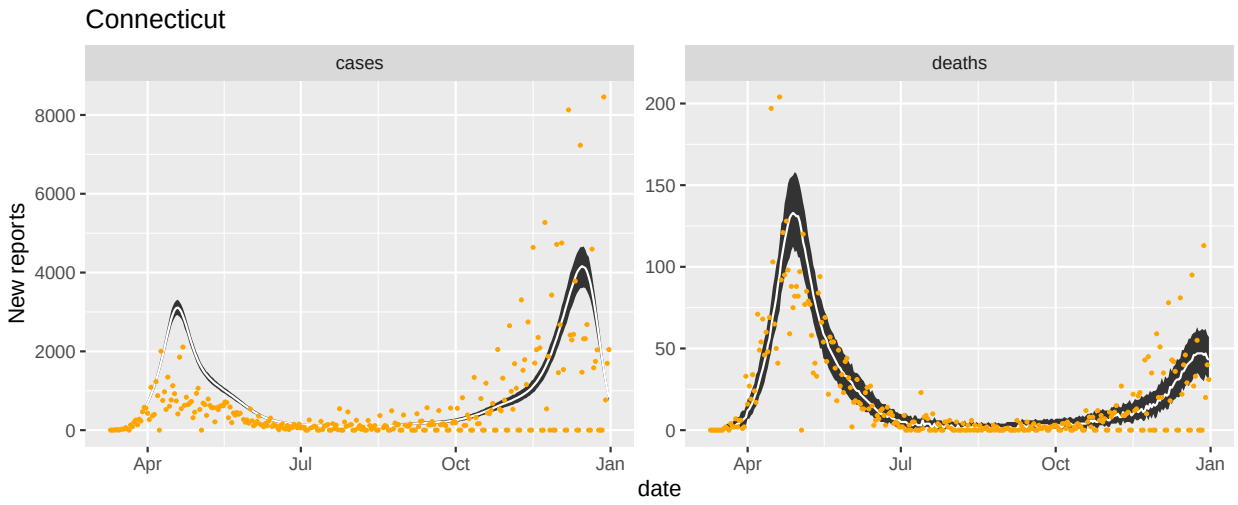
In the following plots, orange points are the reported daily cases or deaths, white line shows the median of the smoothed filtering distributions, and the grey ribbon bounds the 95% prediction interval.

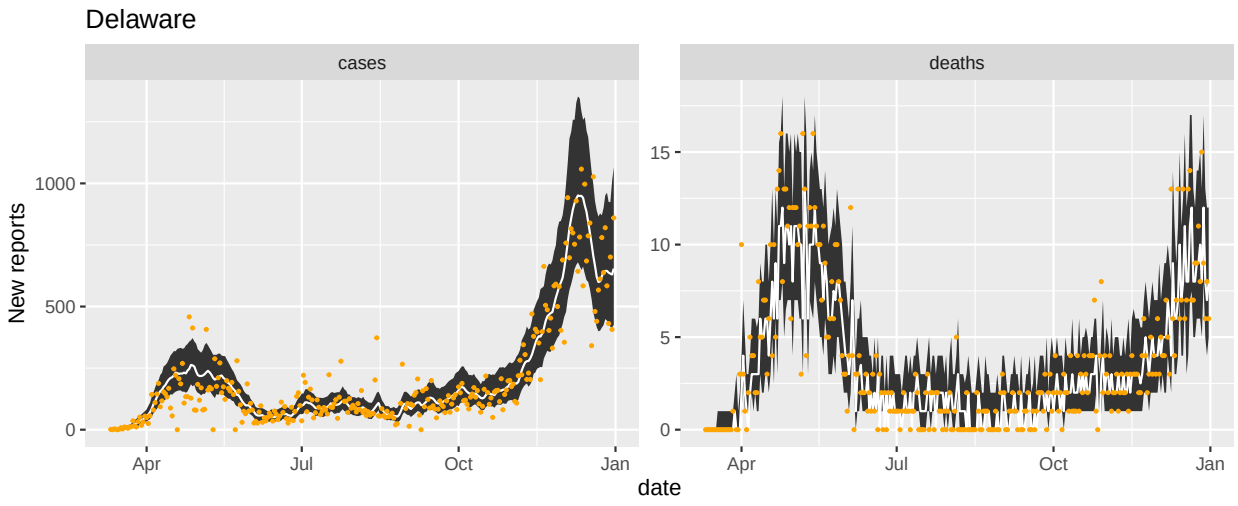


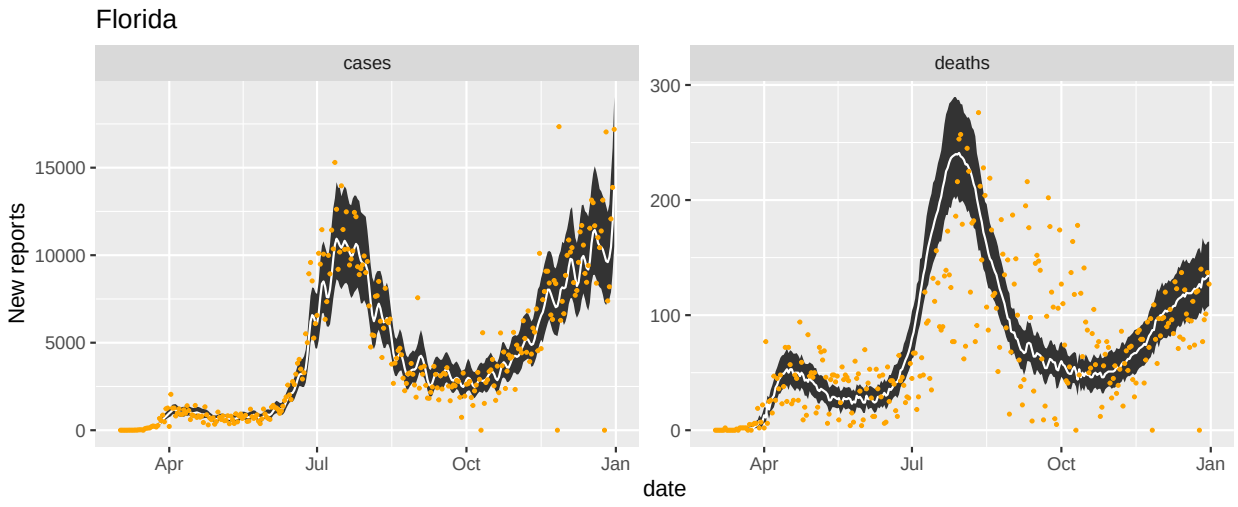

Arkansas
cases
deaths
New reports
4000
3000
2000
1000
0
150
100
50
0
Apr
Jul
Oct
Jan
date

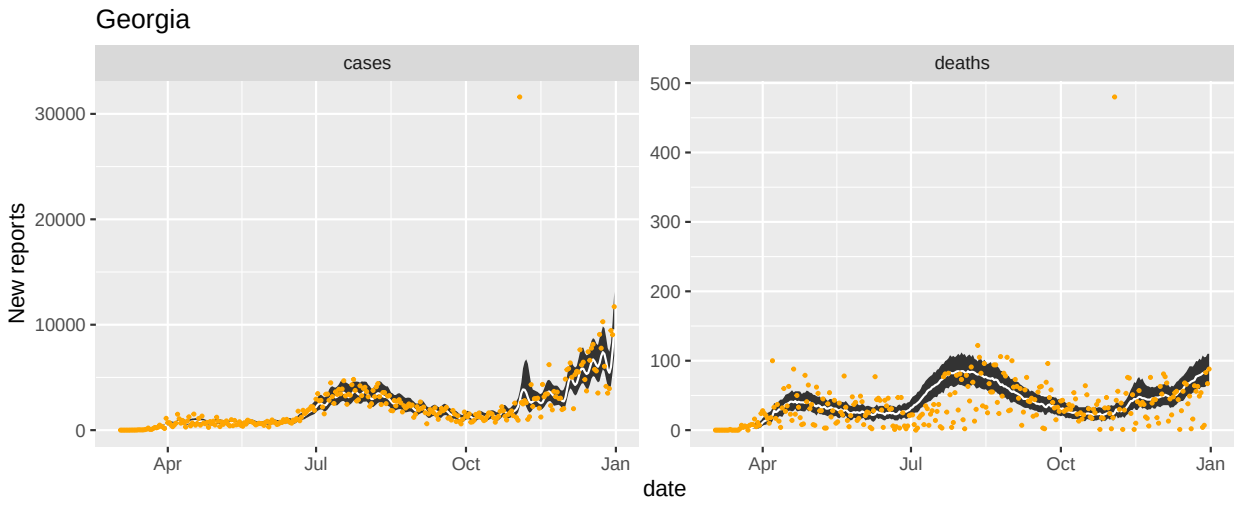

California
cases
deaths
New reports
60000
40000
20000
0
600
400
200
0
Apr
Jul
Oct
Jan
date

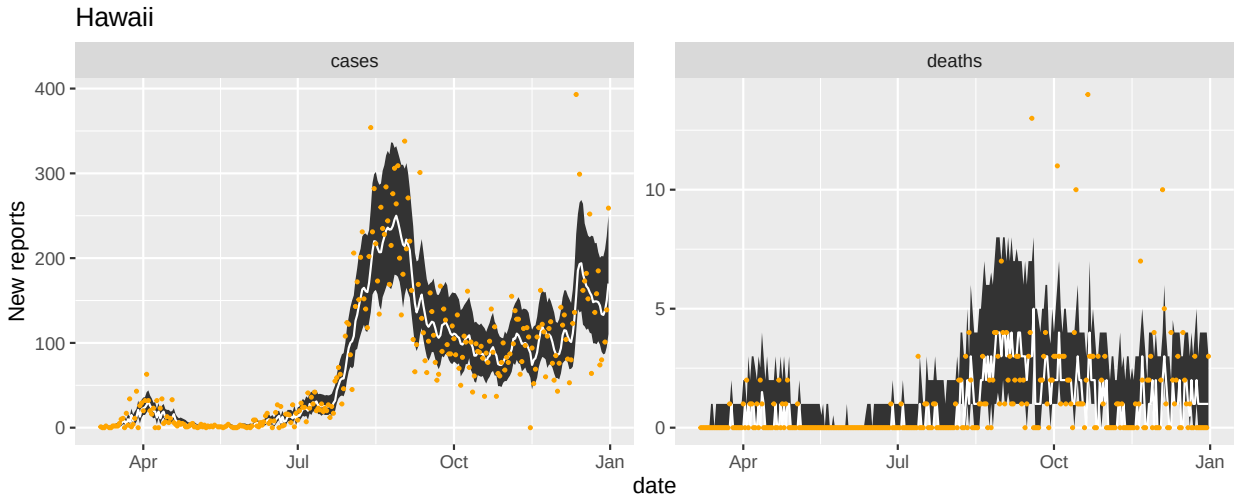

Colorado
cases
deaths
New reports
6000
4000
2000
0
200
100
0
Apr
Jul
Oct
Jan
date

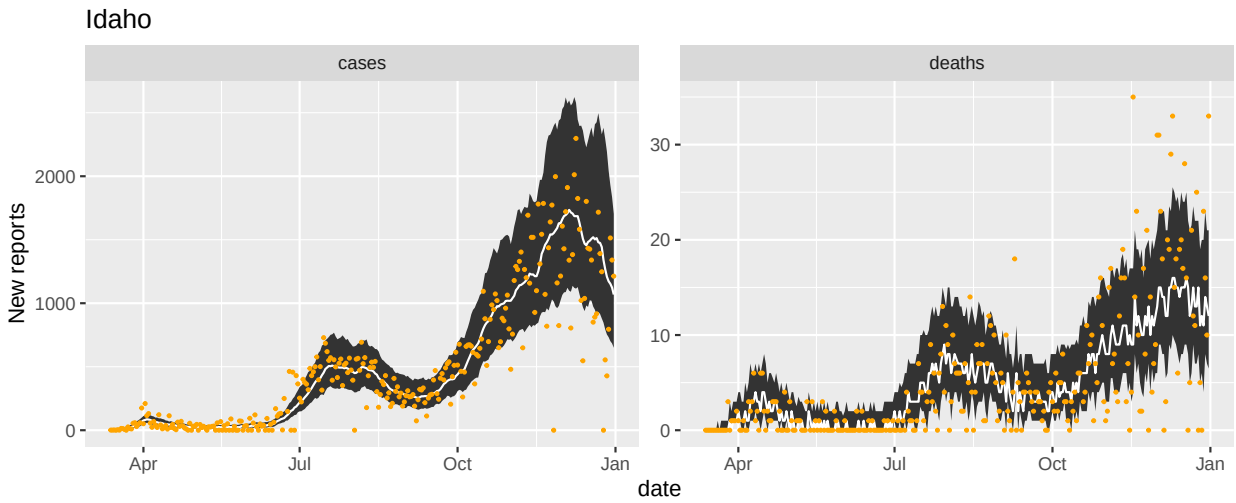

Connecticut
cases
deaths
New reports
8000
6000
4000
2000
0
200
150
100
50
0
Apr
Jul
Oct
Jan
date

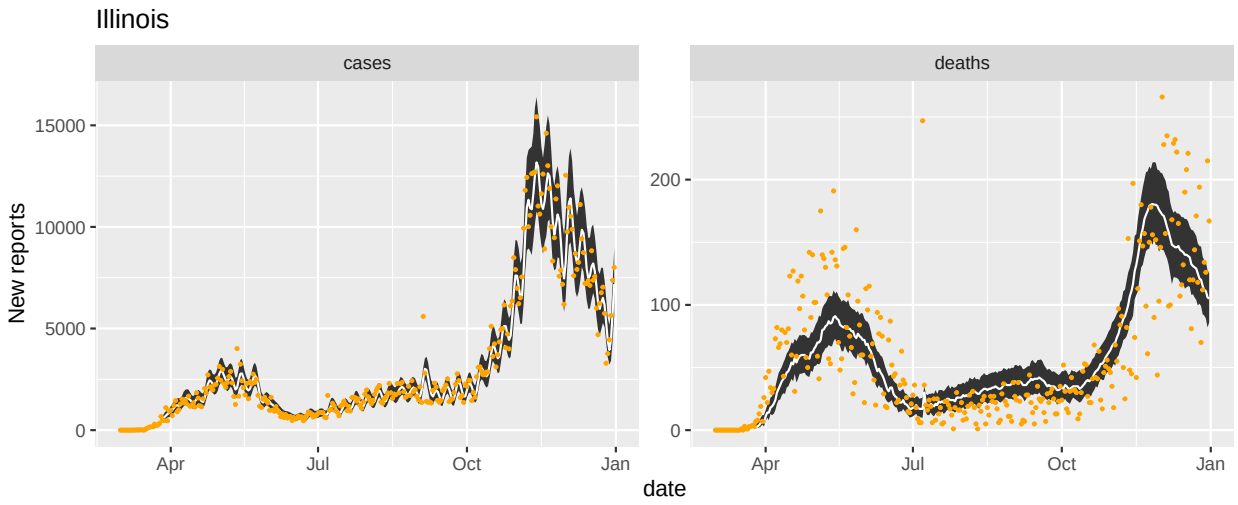

Delaware
cases
deaths
New reports
1000
500
0
15
10
5
0
Apr
Jul
Oct
Jan
date

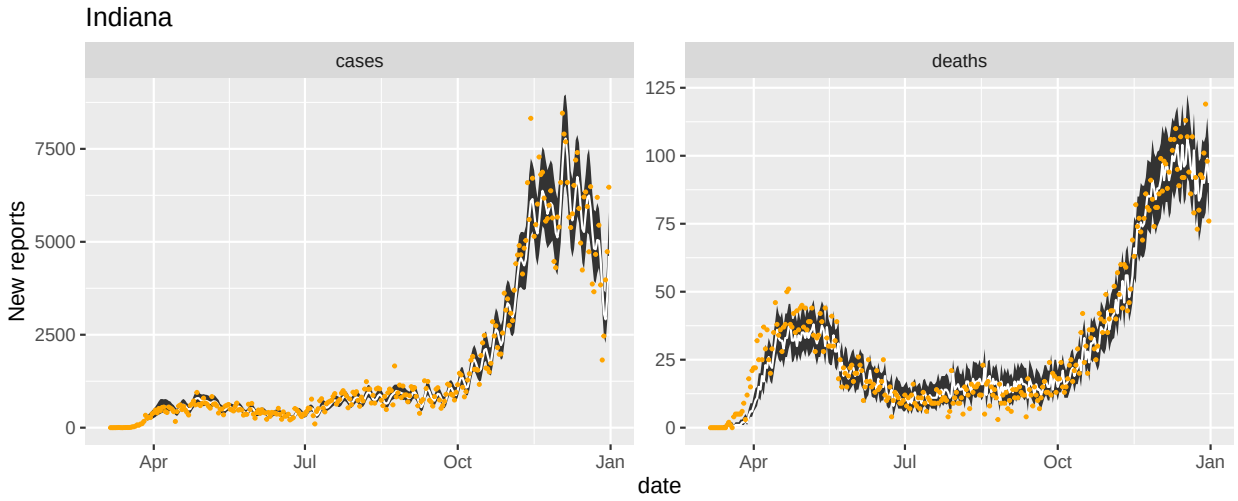

Florida
cases
deaths
New reports
15000
10000
5000
0
300
200
100
0
Apr
Jul
Oct
Jan
date

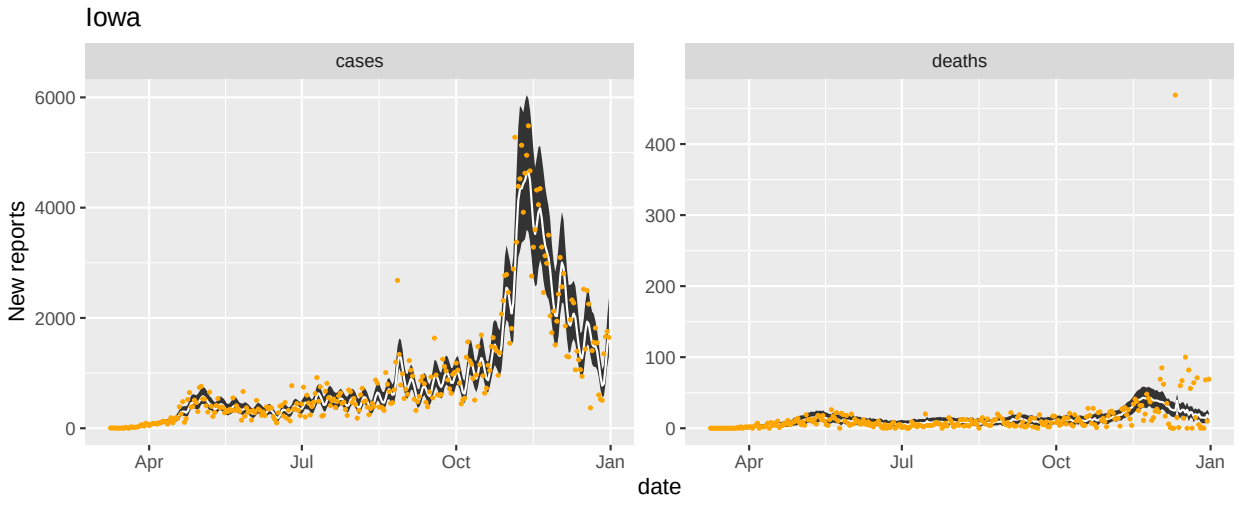

Georgia
cases
deaths
New reports
30000
20000
10000
0
500
400
300
200
100
0
Apr
Jul
Oct
Jan
date

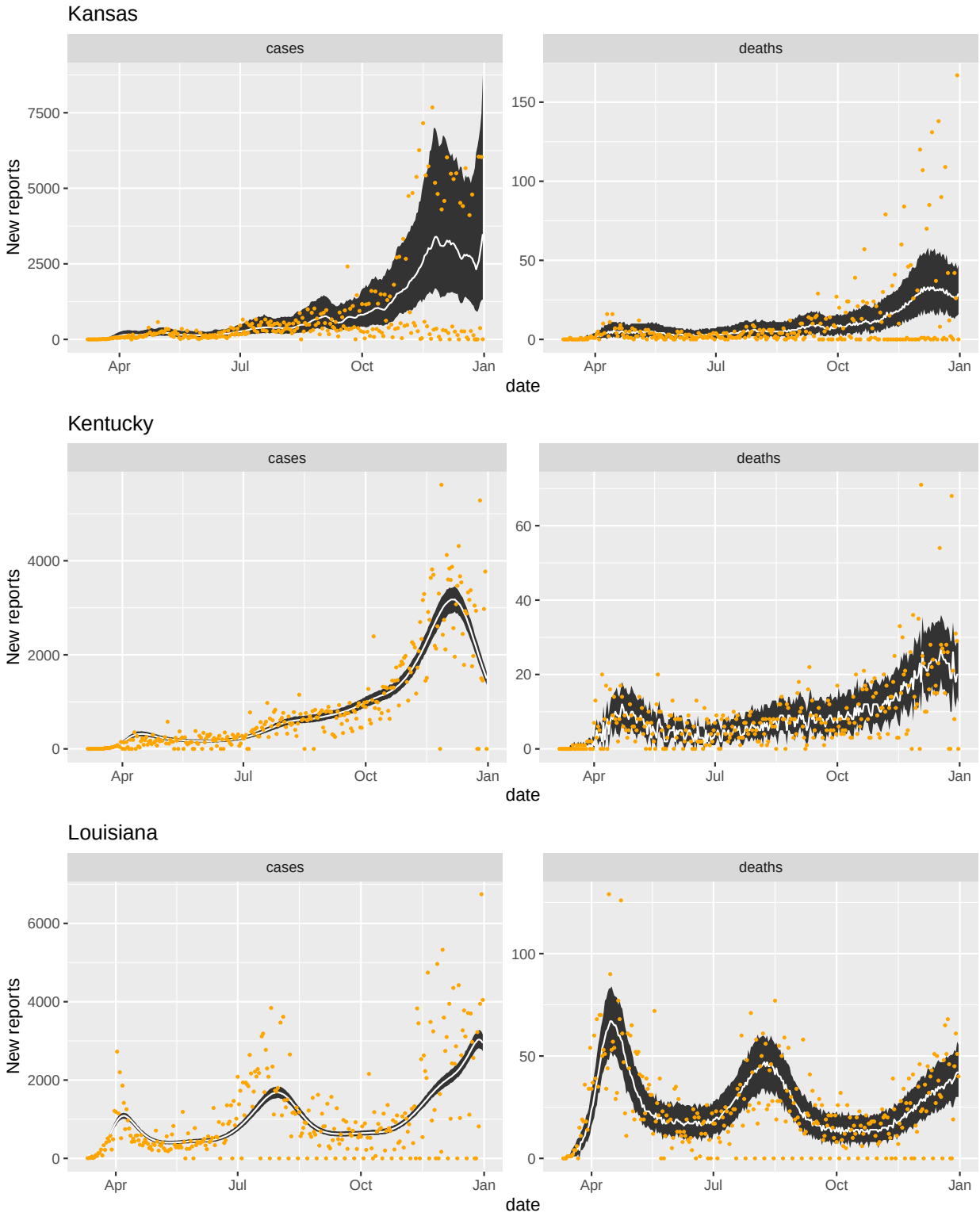

Hawaii
cases
deaths
New reports
400
300
200
100
0
10
5
0
Apr
Jul
Oct
Jan
date

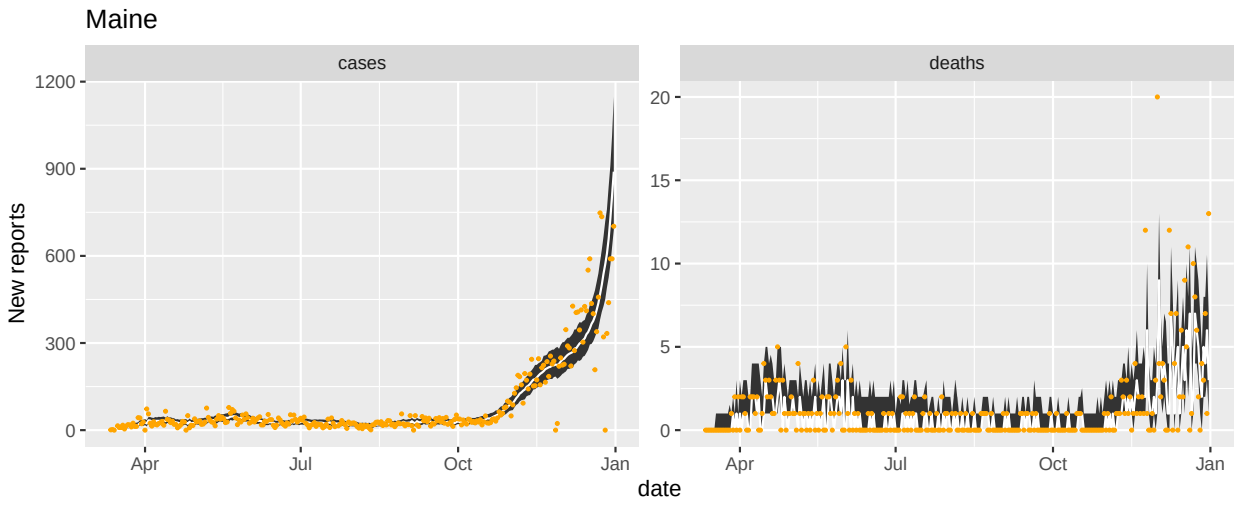

Idaho
cases
deaths
New reports
2000
1000
0
30
20
10
0
Apr
Jul
Oct
Jan
date

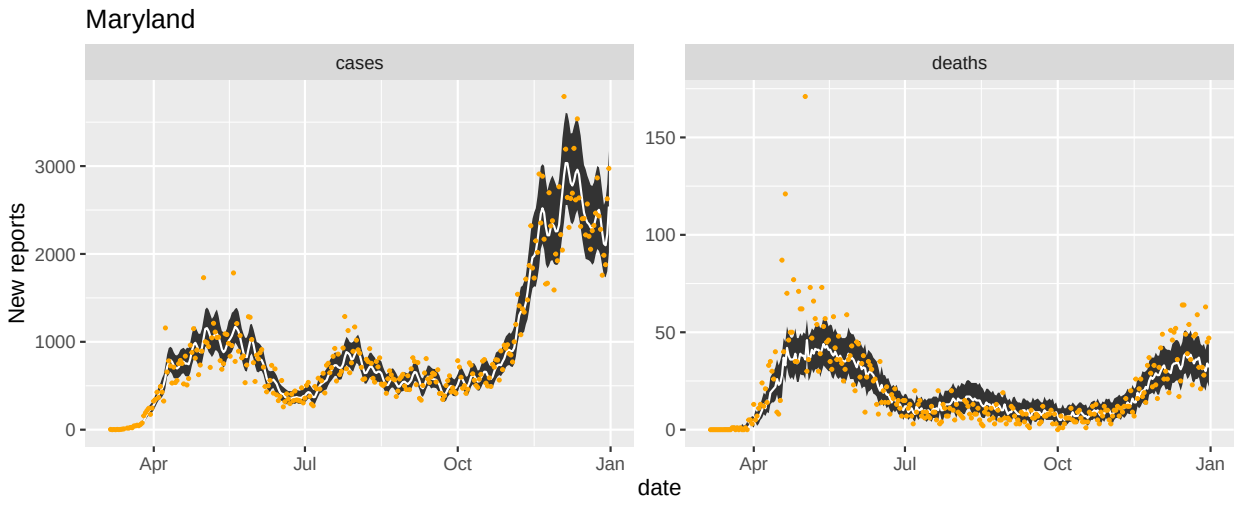

Illinois
cases
deaths
New reports
15000
10000
5000
0
200
100
0
Apr
Jul
Oct
Jan
date

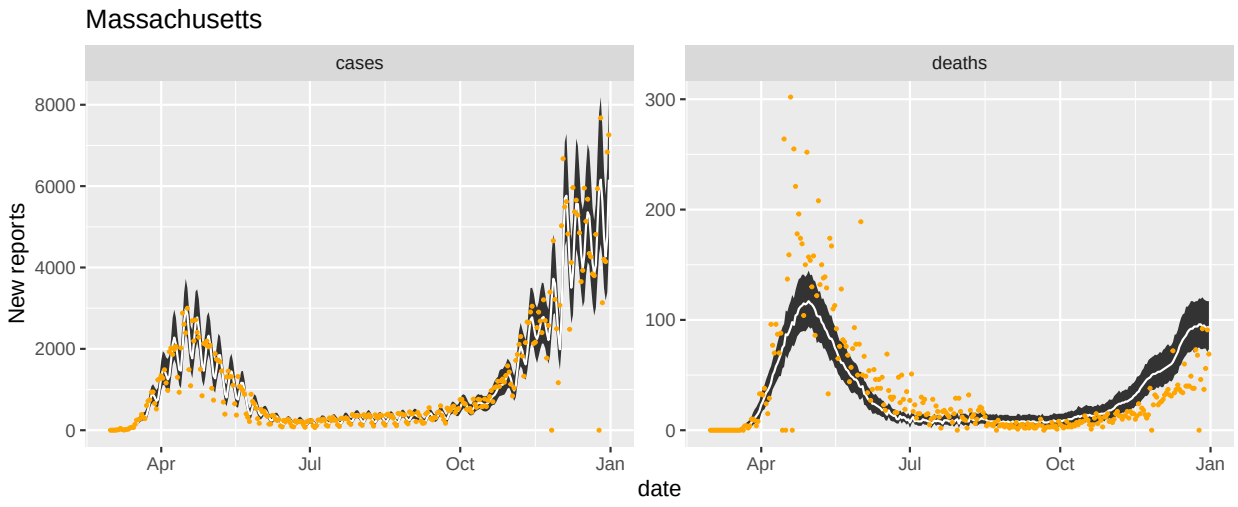

Indiana
cases
deaths
New reports
7500
5000
2500
0
125
100
75
50
25
0
Apr
Jul
Oct
Jan
date

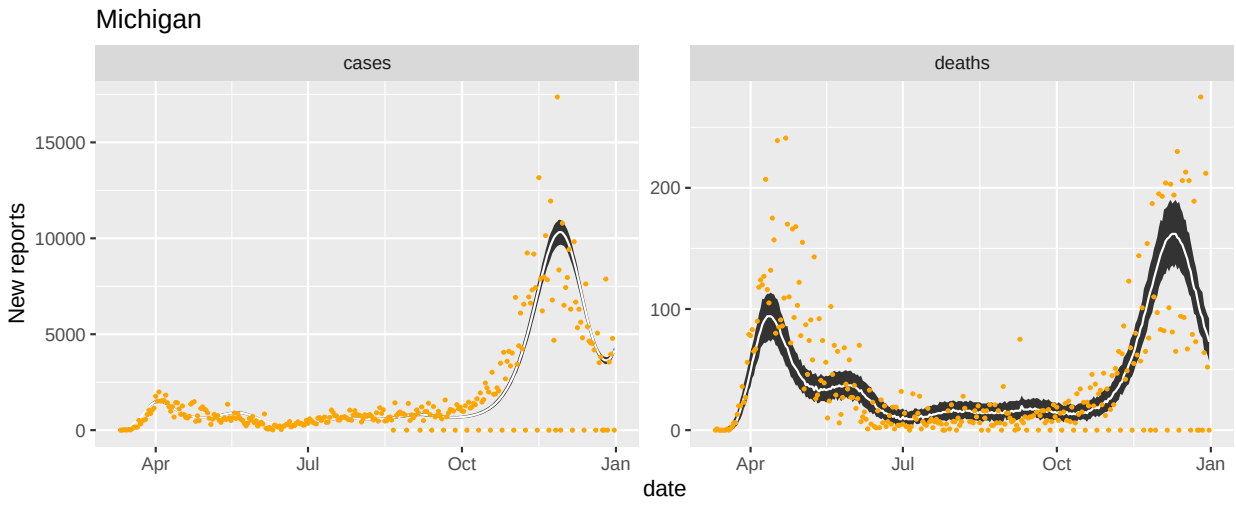

Iowa
cases
deaths
New reports
6000
4000
2000
0
400
300
200
100
0
Apr
Jul
Oct
Jan
date

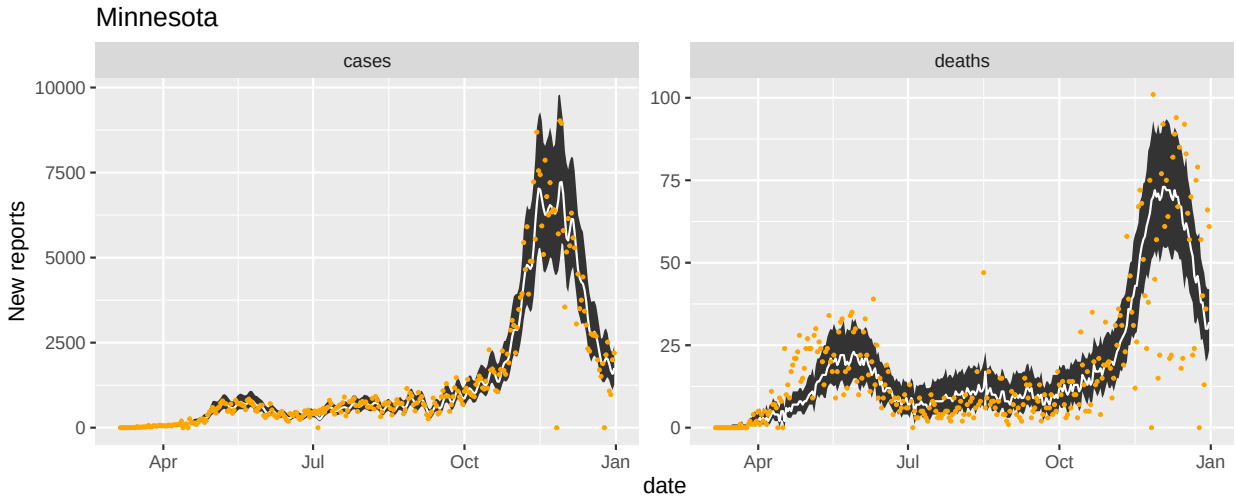

Kansas
cases
deaths
New reports
7500
5000
2500
0
150
100
50
0
Apr
Jul
Oct
Jan
date
Kentucky
cases
deaths
New reports
4000
2000
0
60
40
20
0
Apr
Jul
Oct
Jan
date
Louisiana
cases
deaths
New reports
6000
4000
2000
0
100
50
0
Apr
Jul
Oct
Jan
date

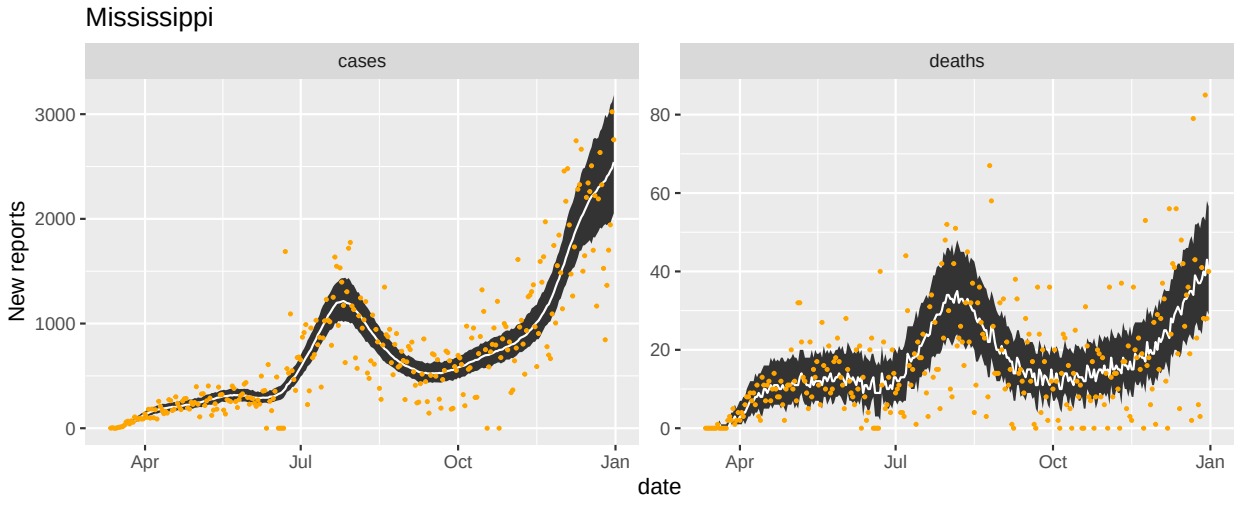

Maine
cases
deaths
New reports
1200
900
600
300
0
20
15
10
5
0
Apr
Jul
Oct
Jan
date

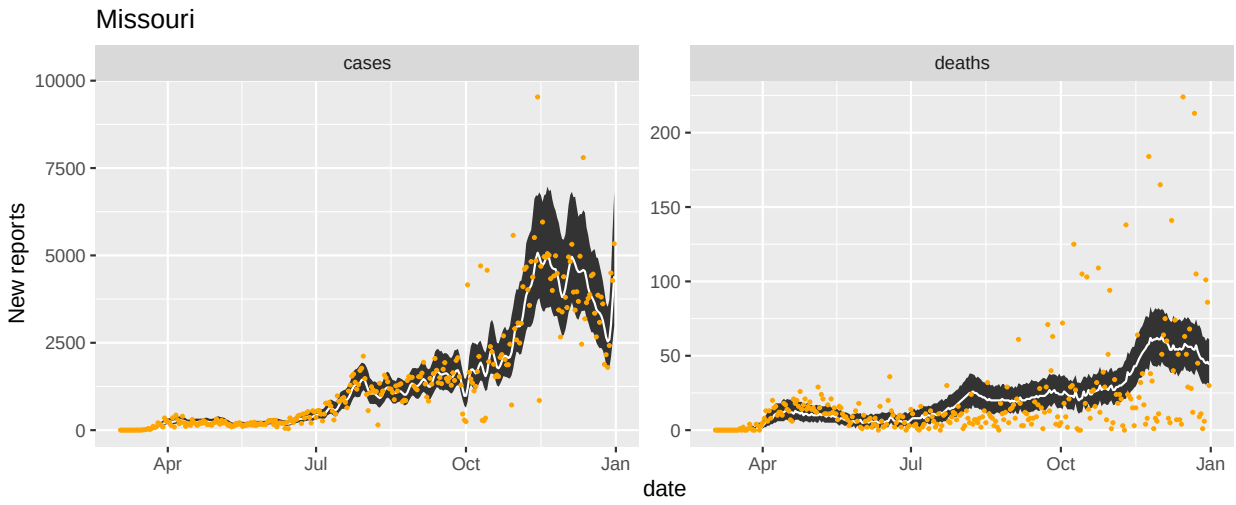

Maryland
cases
deaths
New reports
3000
2000
1000
0
150
100
50
0
Apr
Jul
Oct
Jan
date

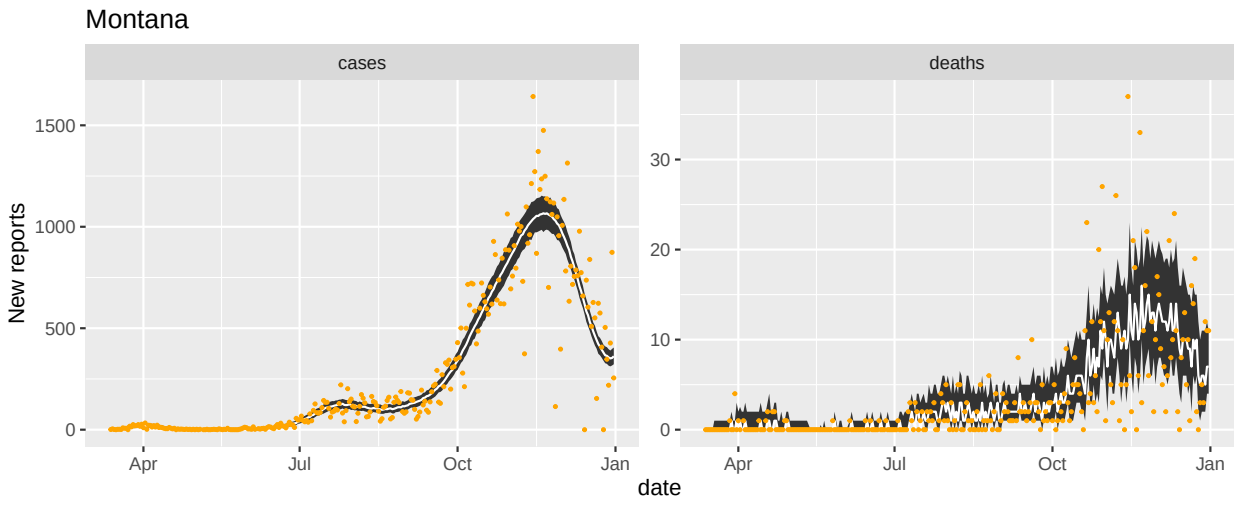

Massachusetts
cases
deaths
New reports
8000
6000
4000
2000
0
300
200
100
0
Apr
Jul
Oct
Jan
date

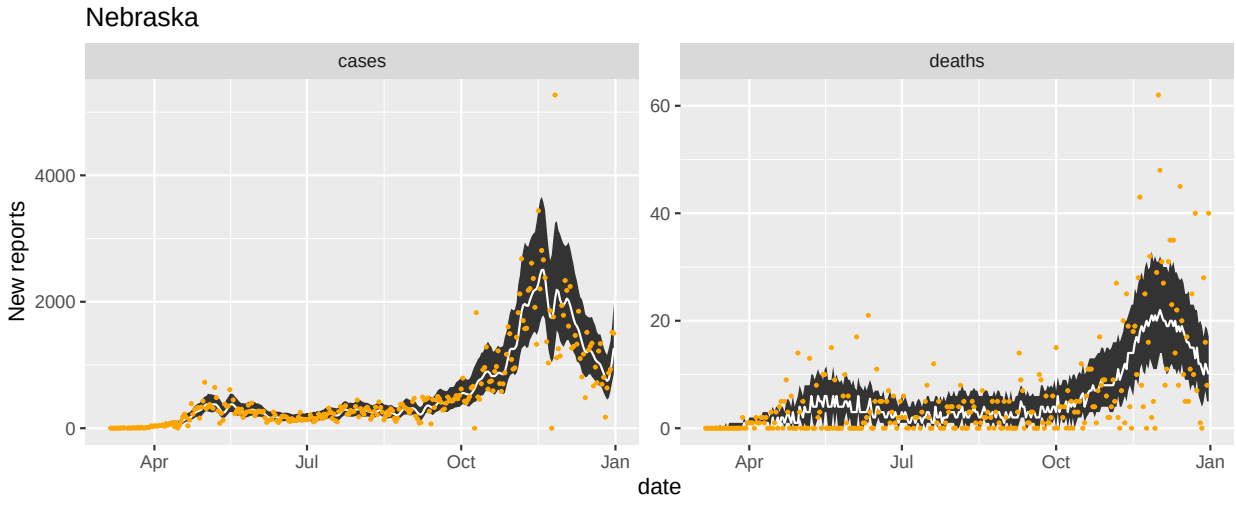

Michigan
cases
deaths
New reports
15000
10000
5000
0
200
100
0
Apr
Jul
Oct
Jan
date

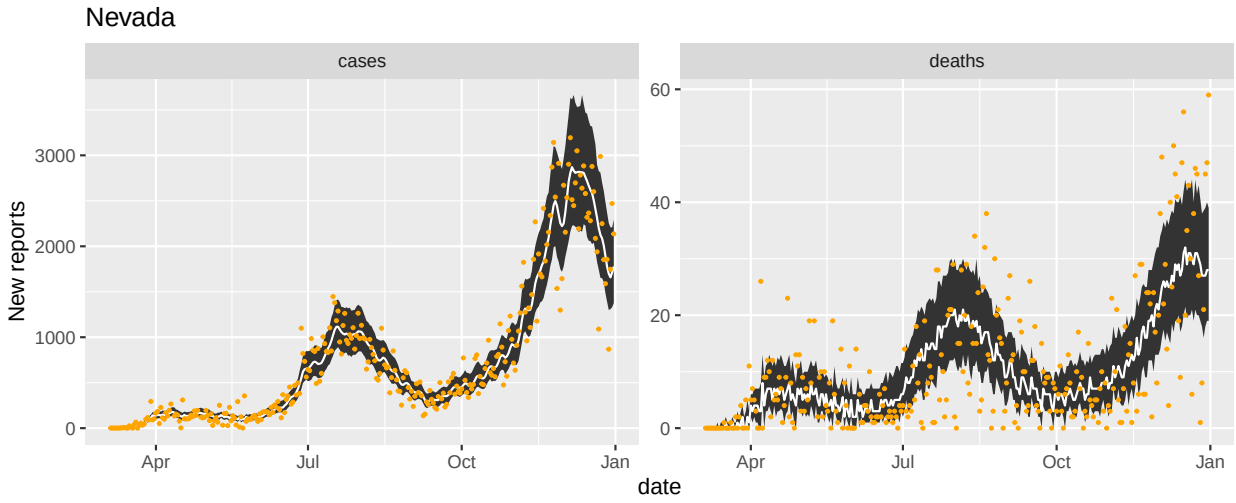

Minnesota
cases
deaths
New reports
10000
7500
5000
2500
0
100
75
50
25
0
Apr
Jul
Oct
Jan
date

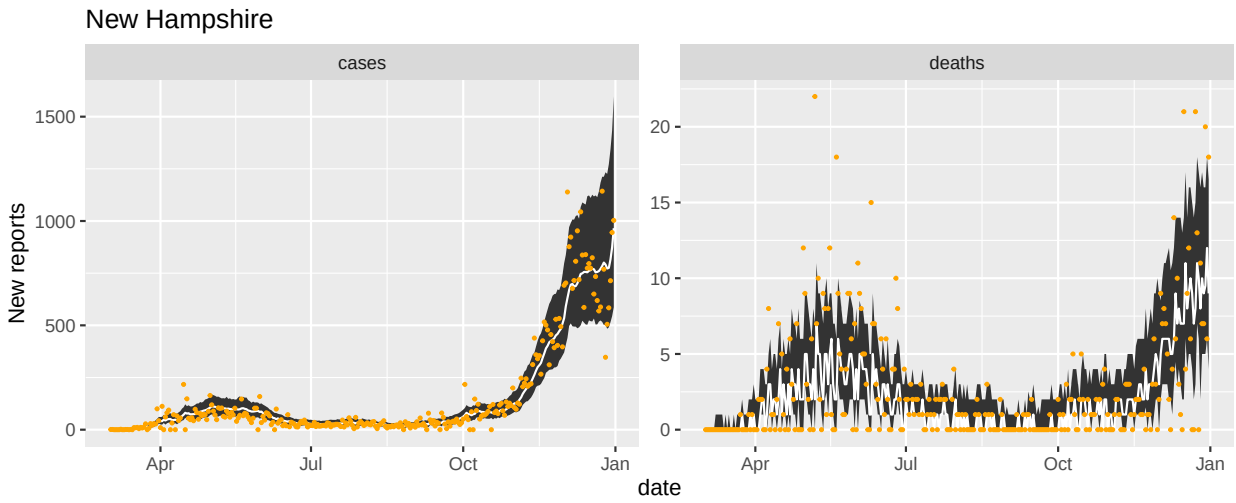

Mississippi
cases
deaths
New reports
3000
2000
1000
0
80
60
40
20
0
Apr
Jul
Oct
Jan
date

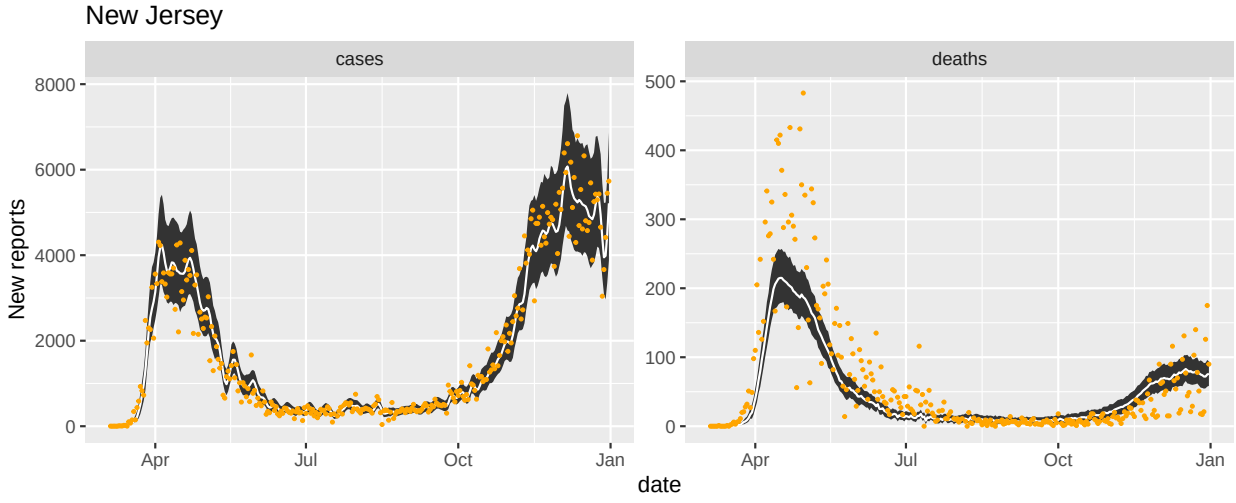

Missouri
cases
deaths
New reports
10000
7500
5000
2500
0
200
150
100
50
0
Apr
Jul
Oct
Jan
date

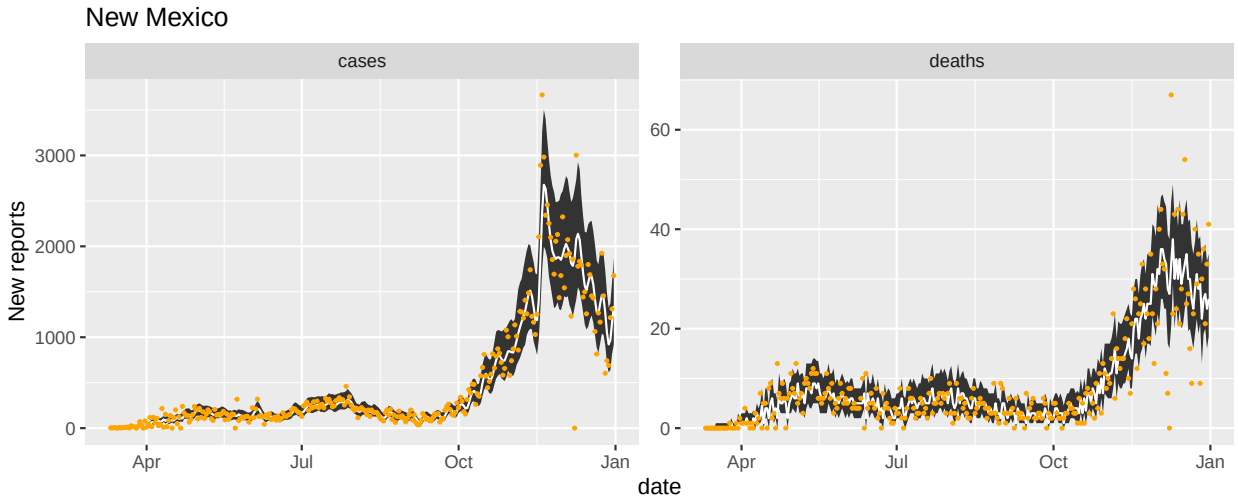

Montana
cases
deaths
New reports
1500
1000
500
0
30
20
10
0
Apr
Jul
Oct
Jan
date


Nebraska
cases
deaths
New reports
4000
2000
0
60
40
20
0
Apr
Jul
Oct
Jan
date


Nevada
cases
deaths
New reports
3000
2000
1000
0
60
40
20
0
Apr
Jul
Oct
Jan
date

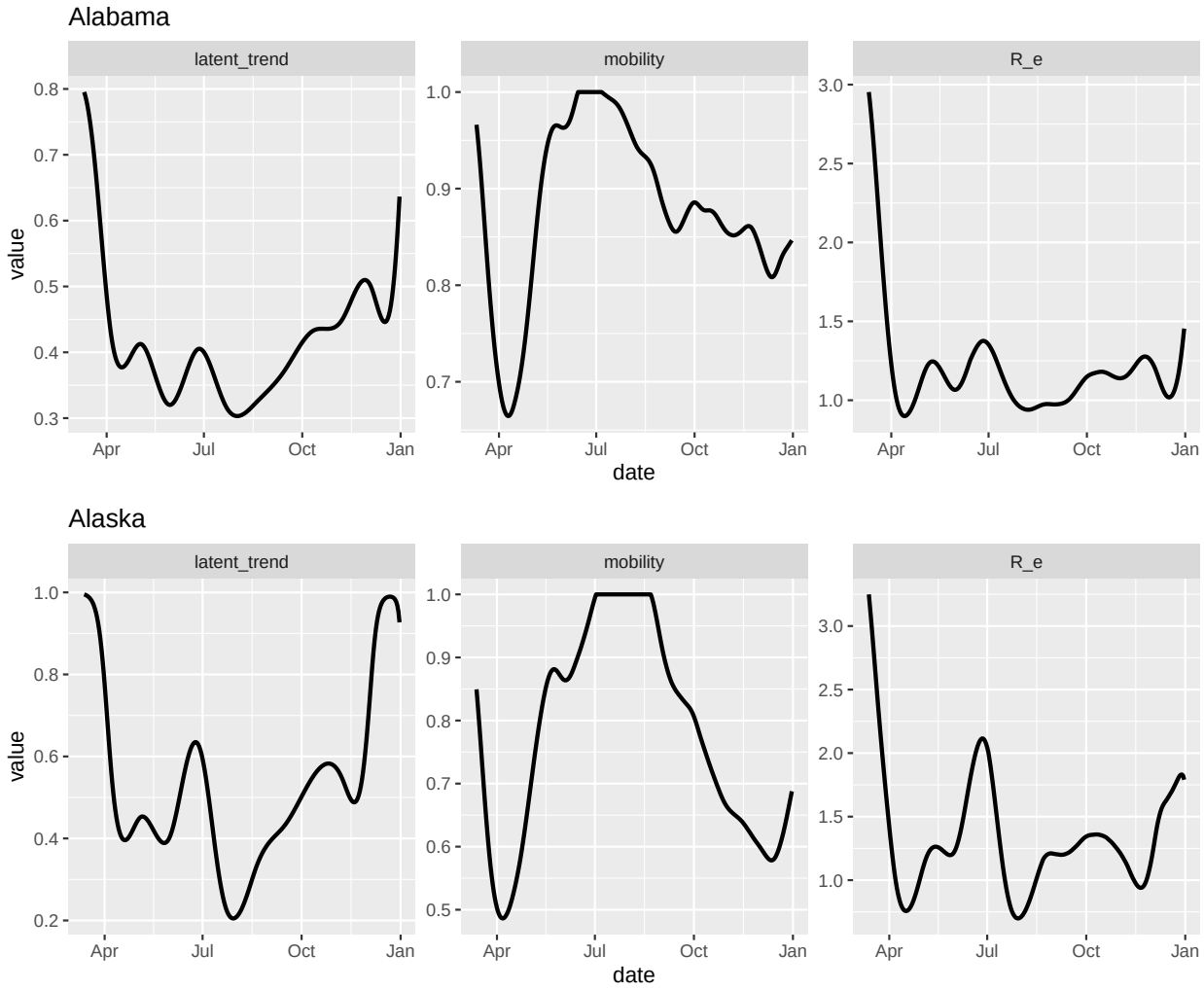

New Hampshire
cases
deaths
New reports
1500
1000
500
0
20
15
10
5
0
Apr
Jul
Oct
Jan
date

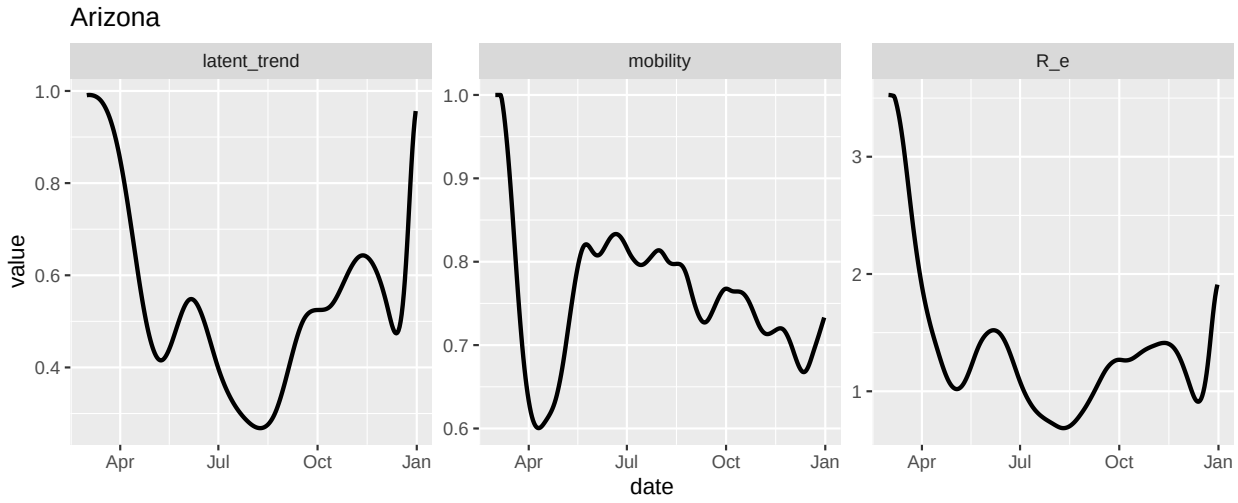

New Jersey
cases
deaths
New reports
8000
6000
4000
2000
0
500
400
300
200
100
0
Apr
Jul
Oct
Jan
date

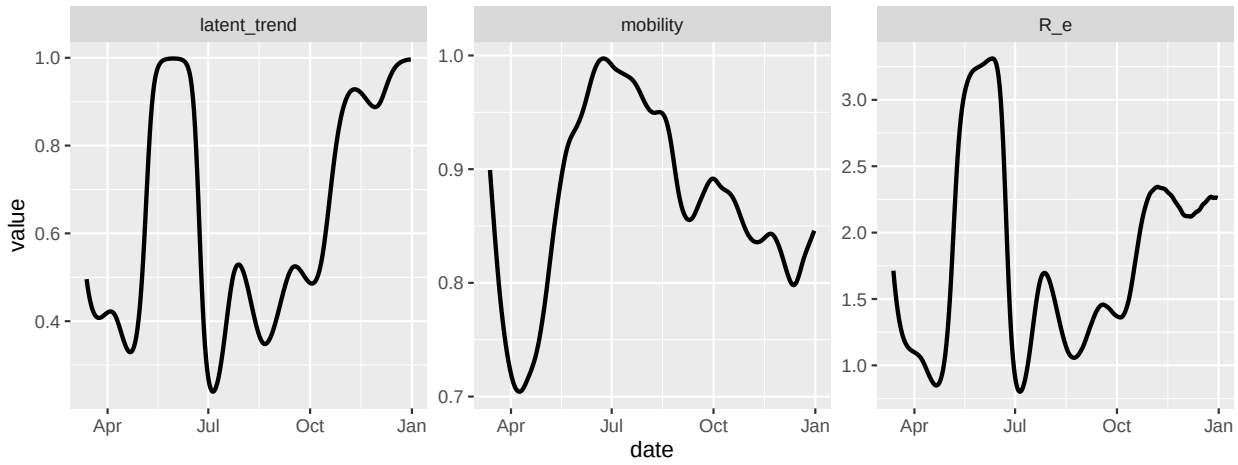

New Mexico
cases
deaths
New reports
3000
2000
1000
0
60
40
20
0
Apr
Jul
Oct
Jan
date

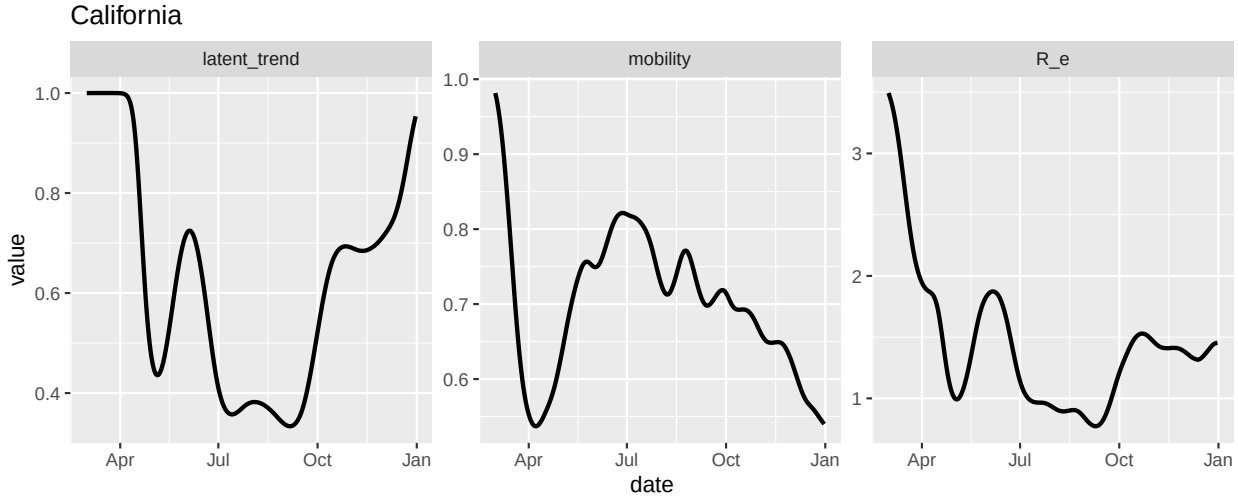
Time series of mobility, estimated latent trend, and $\mathcal{R}_e$

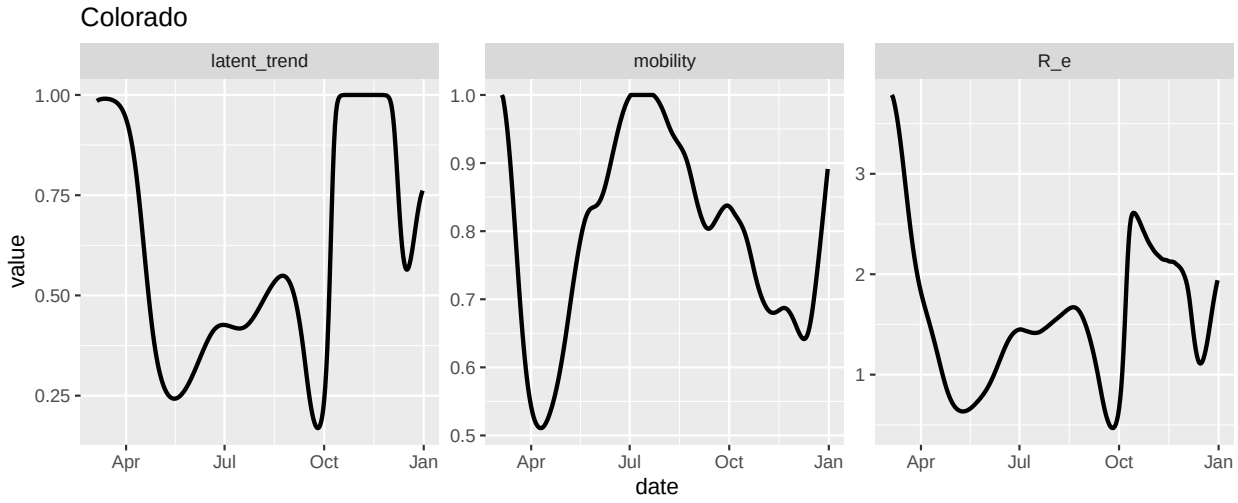
The following plots show the estimated latent trend, relative mobility, and estimated $\mathcal{R}_e$ over time for each state. The latent trend is estimated using the maximum likelihood parameter estimates for the g_i coefficients of the B-spline (see Materials and Methods in main text).

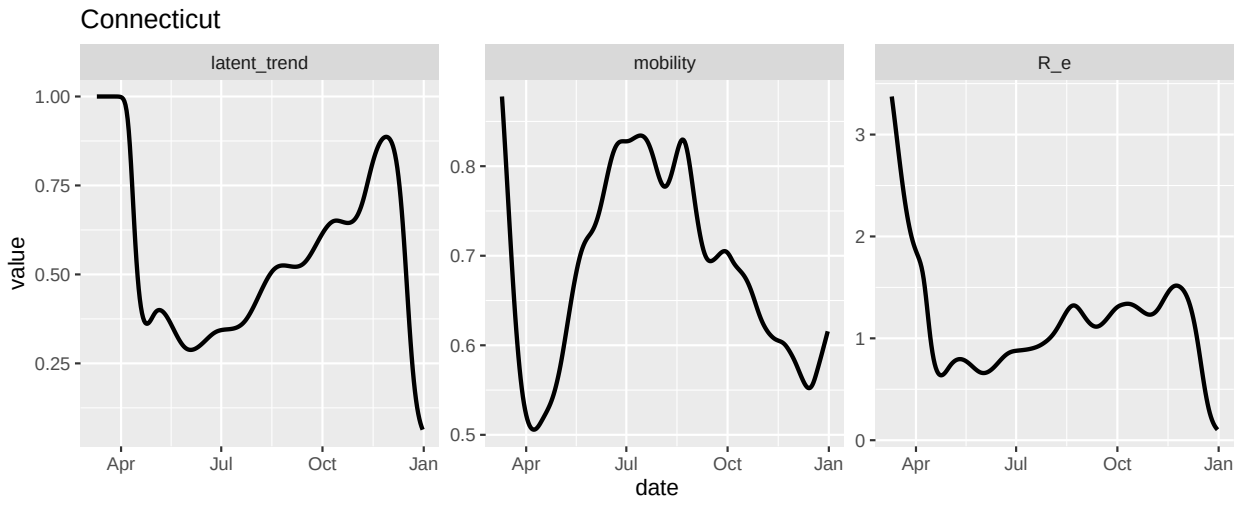


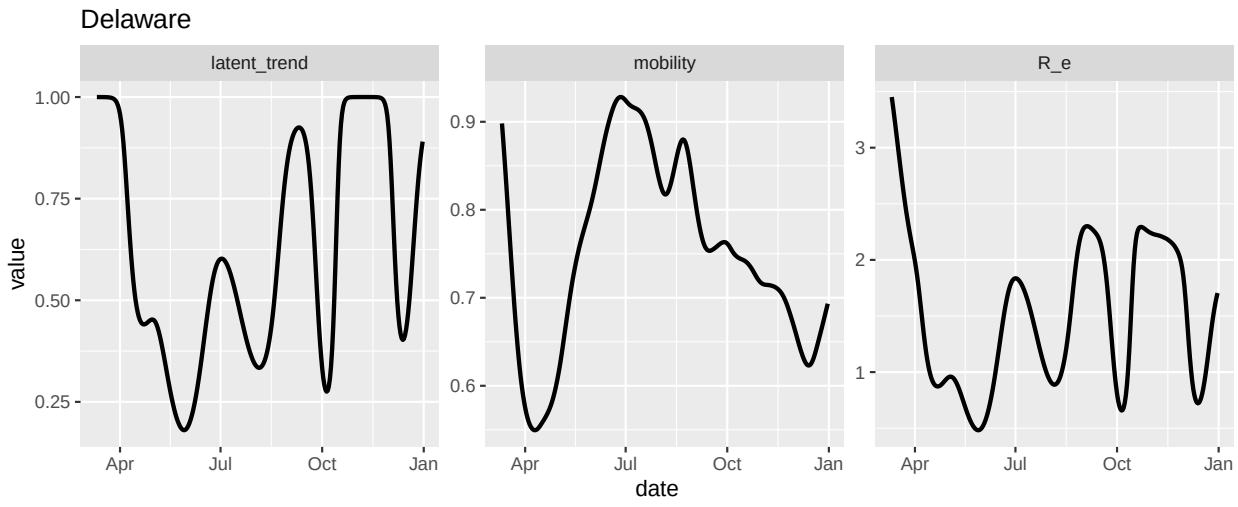

Arizona
latent_trend
mobility
R_e
value
1.0
0.8
0.6
0.4
0.9
0.7
3
2
1
Apr
Jul
Oct
Jan
date

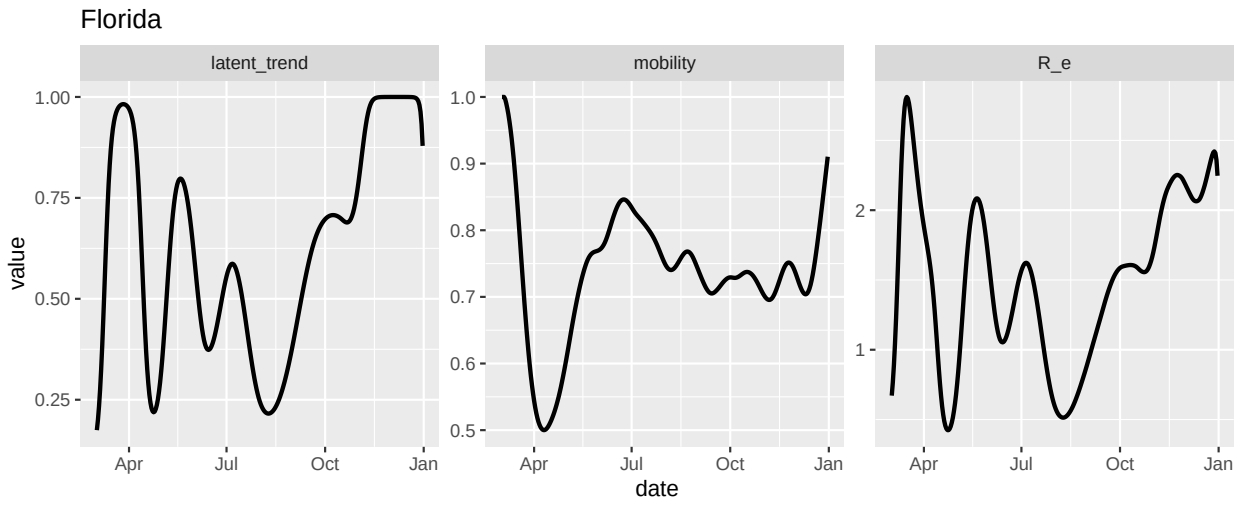

Arkansas
latent_trend
mobility
R_e
value
1.0
0.8
0.6
0.4
0.9
0.7
3.0
2.5
2.0
1.5
1.0
Apr
Jul
Oct
Jan
date

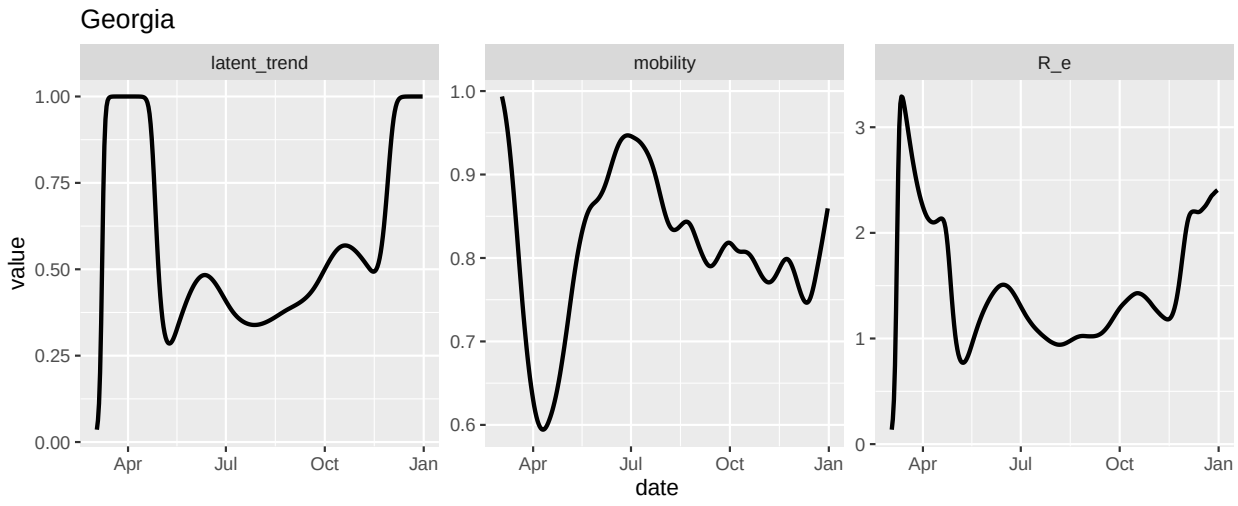

California
latent_trend
mobility
R_e
value
1.0
0.8
0.6
0.4
0.9
0.7
3
2
1
Apr
Jul
Oct
Jan
date

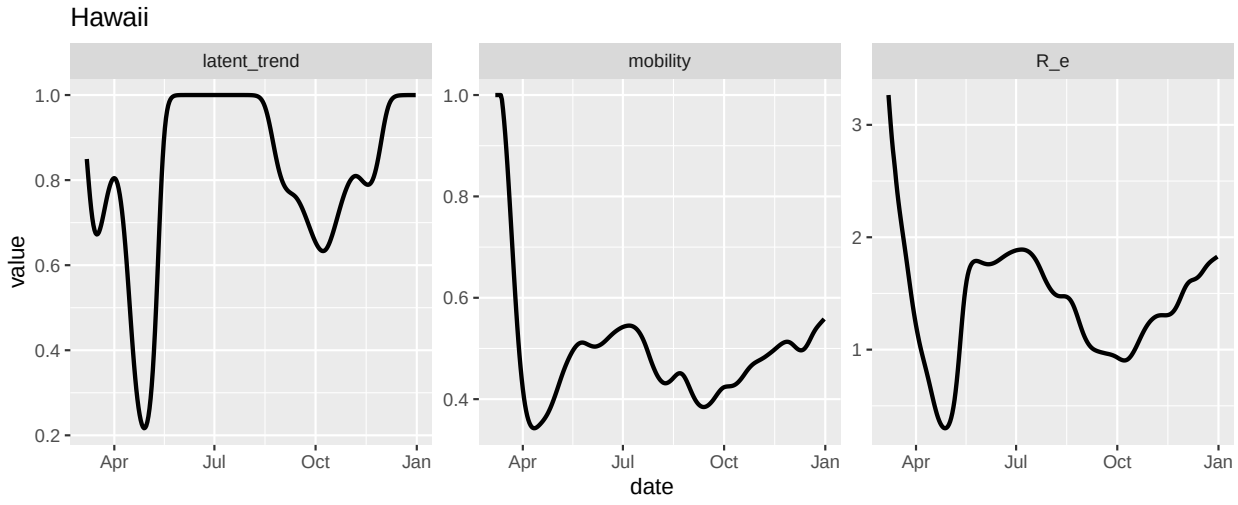

Colorado
latent_trend
mobility
R_e
value
1.00
0.75
0.50
0.25
1.0
0.9
0.8
0.7
0.6
0.5
3
2
1
Apr
Jul
Oct
Jan
date

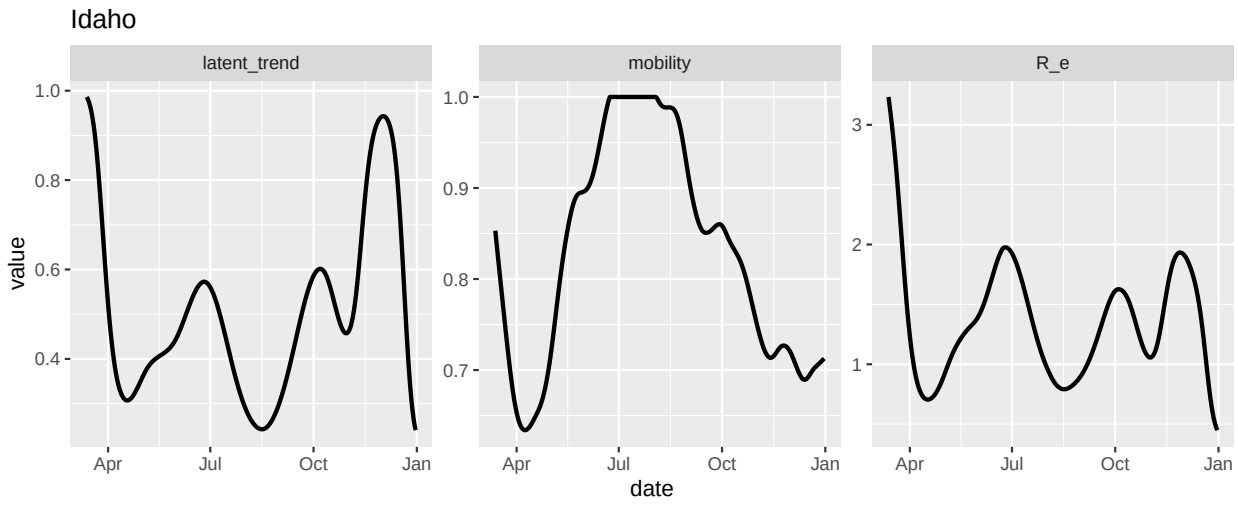

Connecticut
latent_trend
mobility
R_e
value
1.00
0.75
0.50
0.25
0.8
0.7
0.6
0.5
3
2
1
0
Apr
Jul
Oct
Jan
date

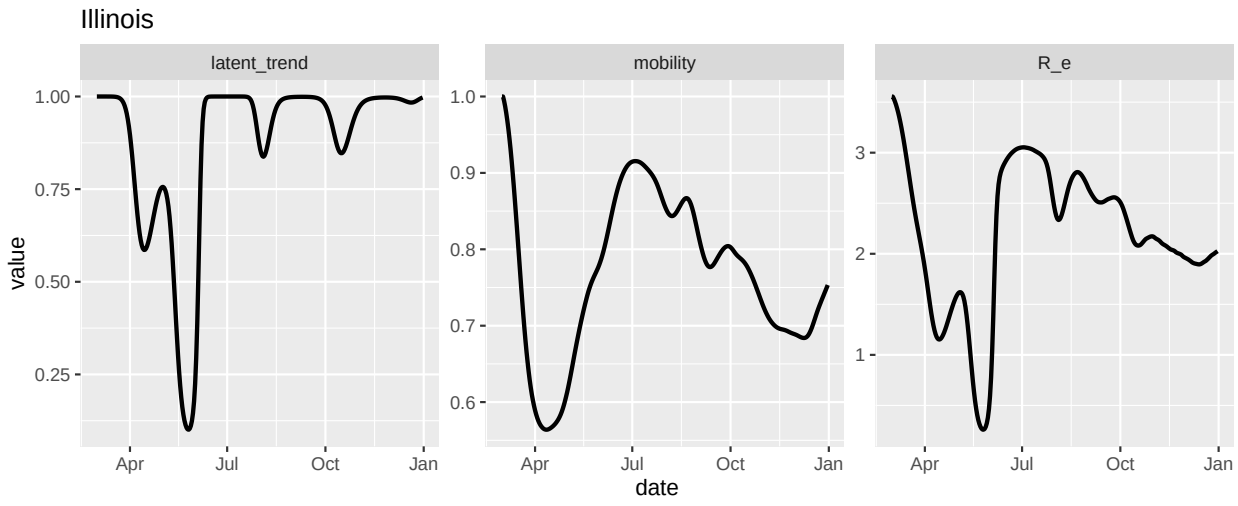

Delaware
latent_trend
mobility
R_e
value
1.00
0.75
0.50
0.25
0.9
0.8
0.7
0.6
3
2
1
Apr
Jul
Oct
Jan
date

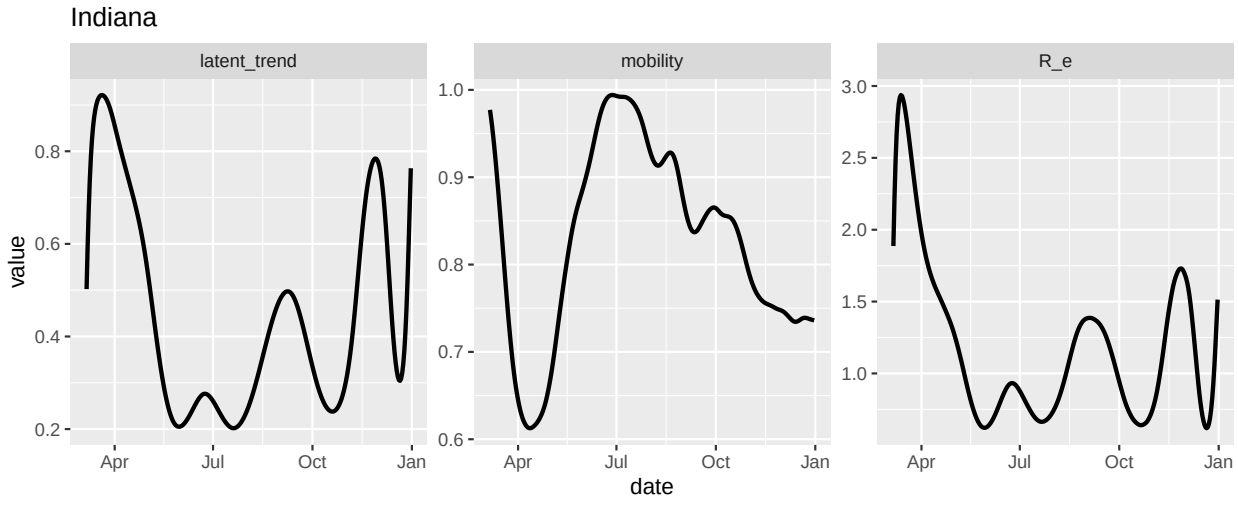

Florida
latent_trend
mobility
R_e
value
1.00
0.75
0.50
0.25
1.0
0.9
0.8
0.7
0.6
0.5
2
1
Apr
Jul
Oct
Jan
date

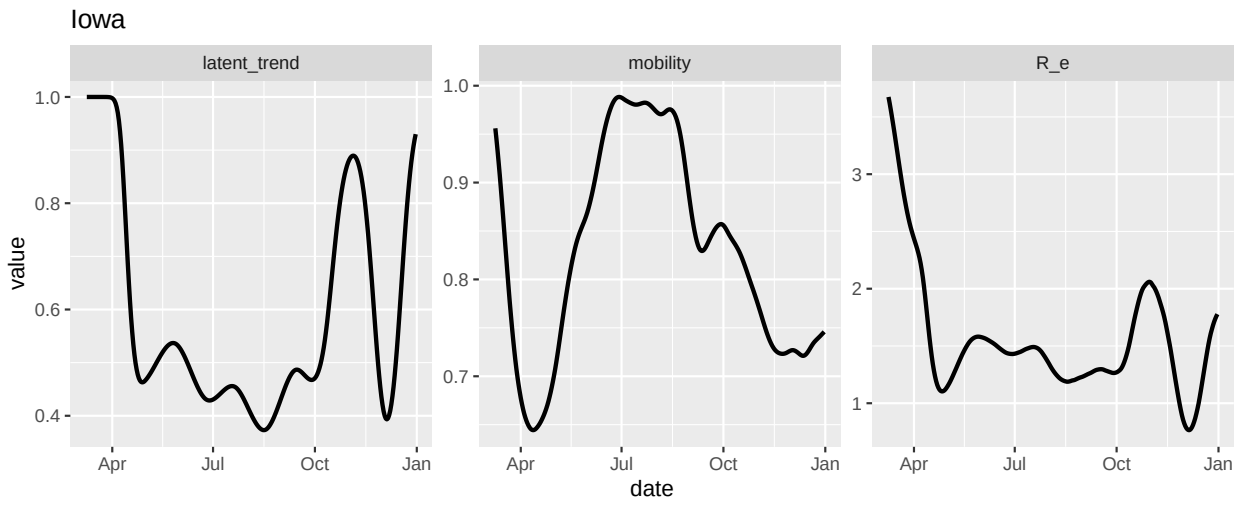

Georgia
latent_trend
mobility
R_e
value
1.00
0.75
0.50
0.25
0.00
1.0
0.9
0.8
0.7
0.6
3
2
1
0
Apr
Jul
Oct
Jan
date

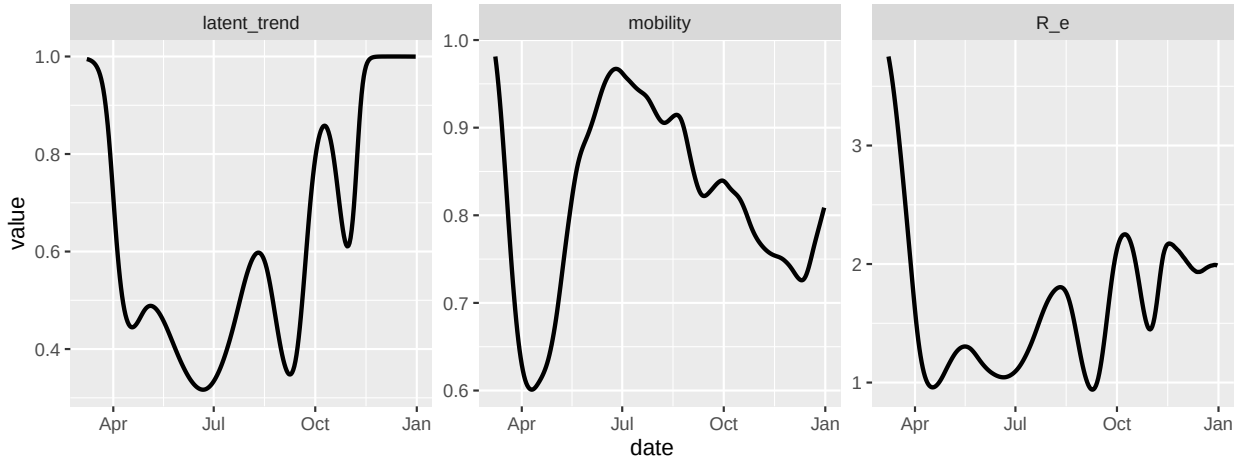

Hawaii
latent_trend
mobility
R_e
value
1.0
0.8
0.6
0.4
0.2
3
2
1
Apr
Jul
Oct
Jan
date

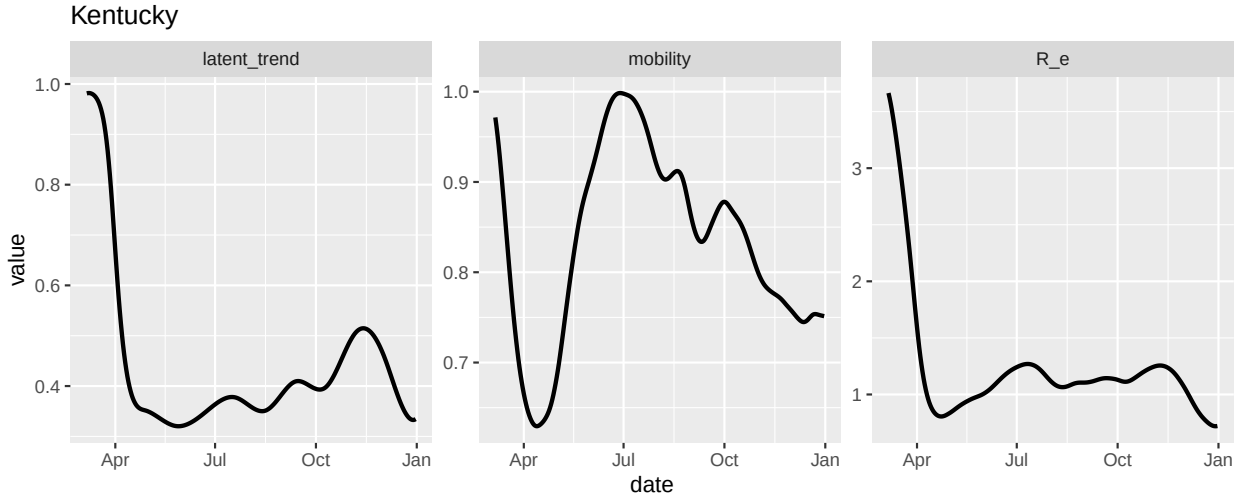

Idaho
latent_trend
mobility
R_e
value
1.0
0.8
0.6
0.4
0.9
0.7
3
2
1
Apr
Jul
Oct
Jan
date

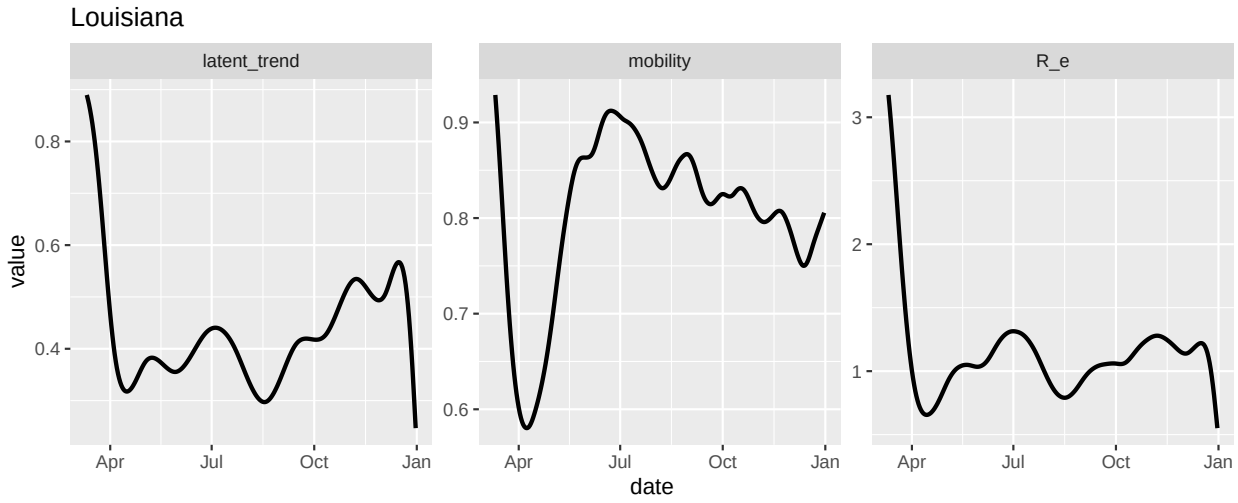

Illinois
latent_trend
mobility
R_e
value
1.00
0.75
0.50
0.25
1.0
0.9
0.8
0.7
0.6
3
2
1
Apr
Jul
Oct
Jan
date

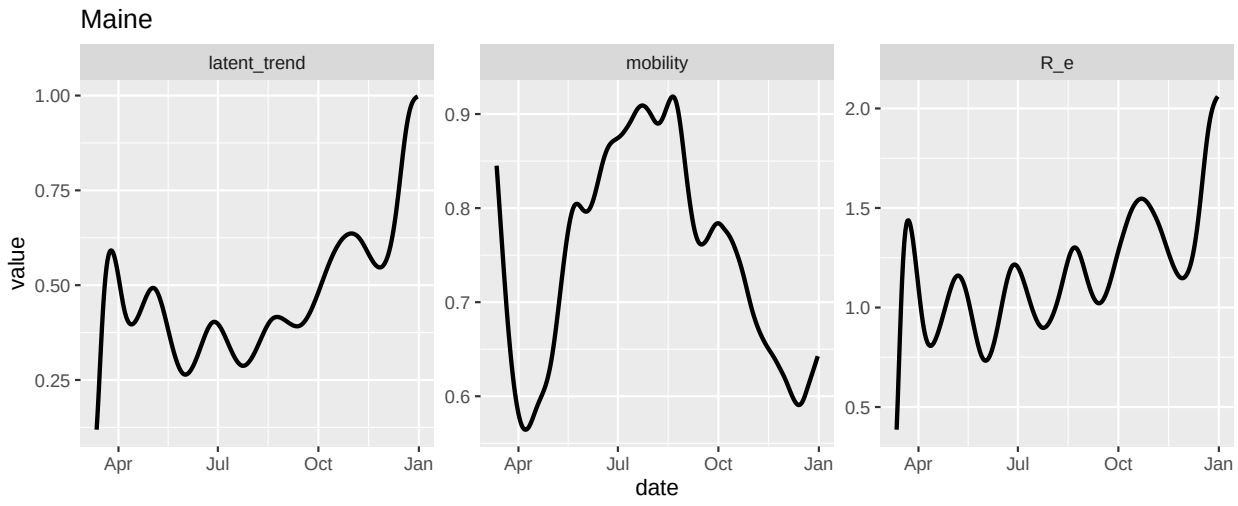

Indiana
latent_trend
mobility
R_e
value
0.8
0.6
0.4
0.2
1.0
0.9
0.8
0.7
0.6
3.0
2.5
2.0
1.5
1.0
Apr
Jul
Oct
Jan
date

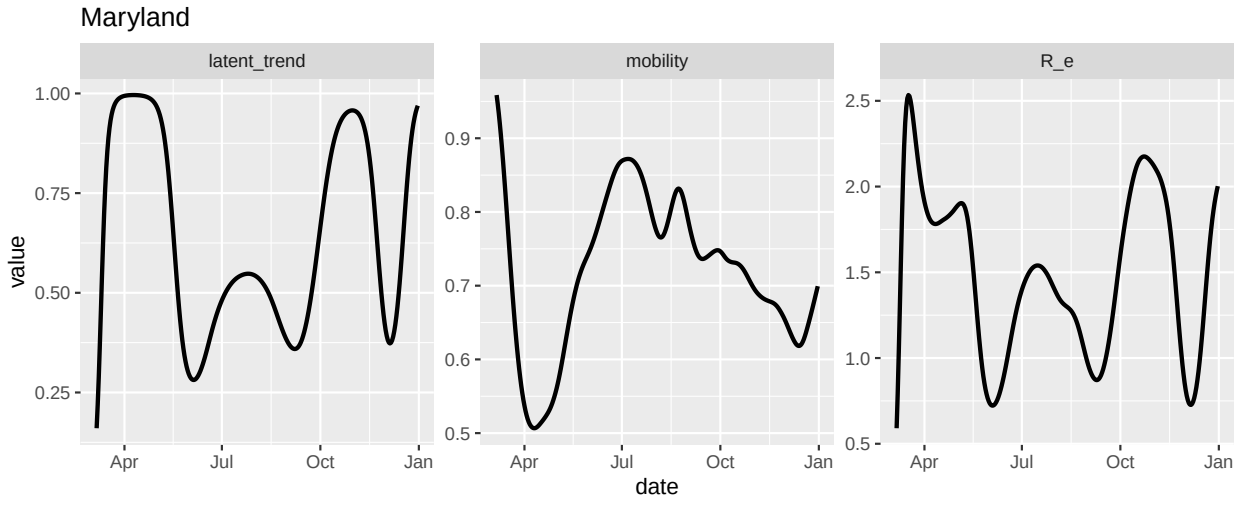

Iowa
latent_trend
mobility
R_e
value
1.0
0.8
0.6
0.4
0.9
0.7
3
2
1
Apr
Jul
Oct
Jan
date

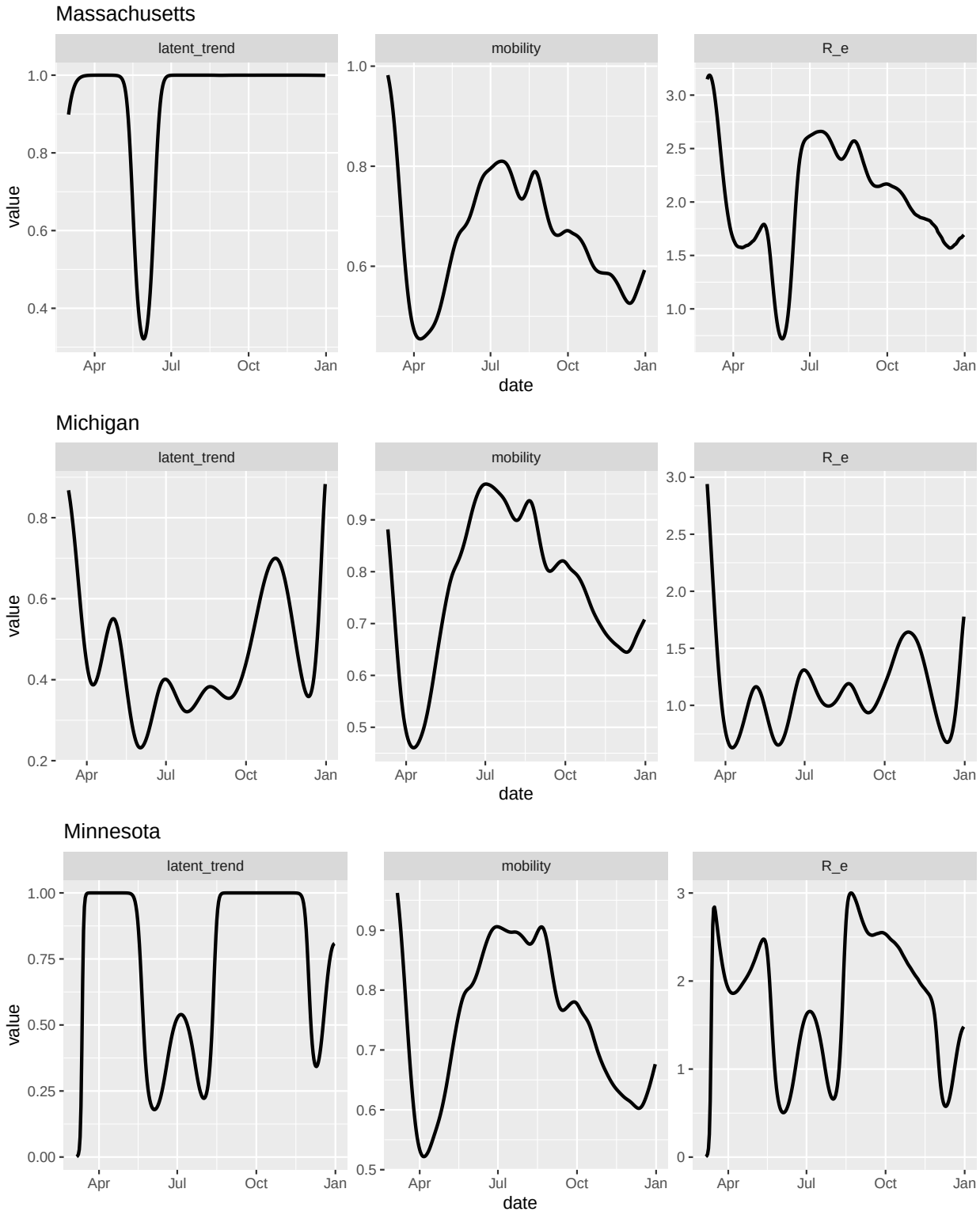

Kansas
latent_trend
mobility
R_e
value
1.0
0.8
0.6
0.4
0.9
0.7
3
2
1
Apr
Jul
Oct
Jan
date

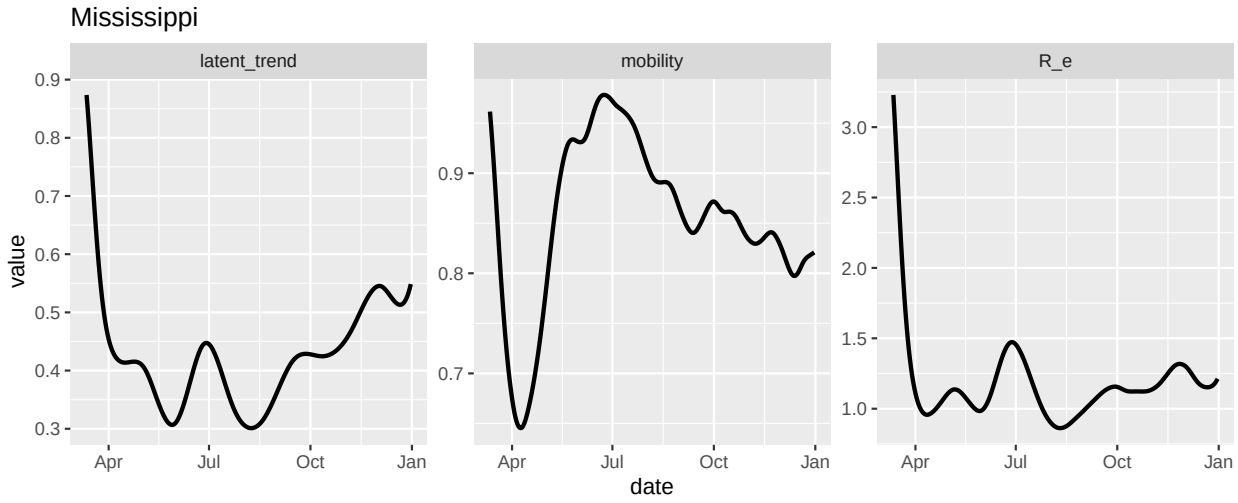

Kentucky
latent_trend
mobility
R_e
value
1.0
0.8
0.6
0.4
0.9
0.7
3
2
1
Apr
Jul
Oct
Jan
date

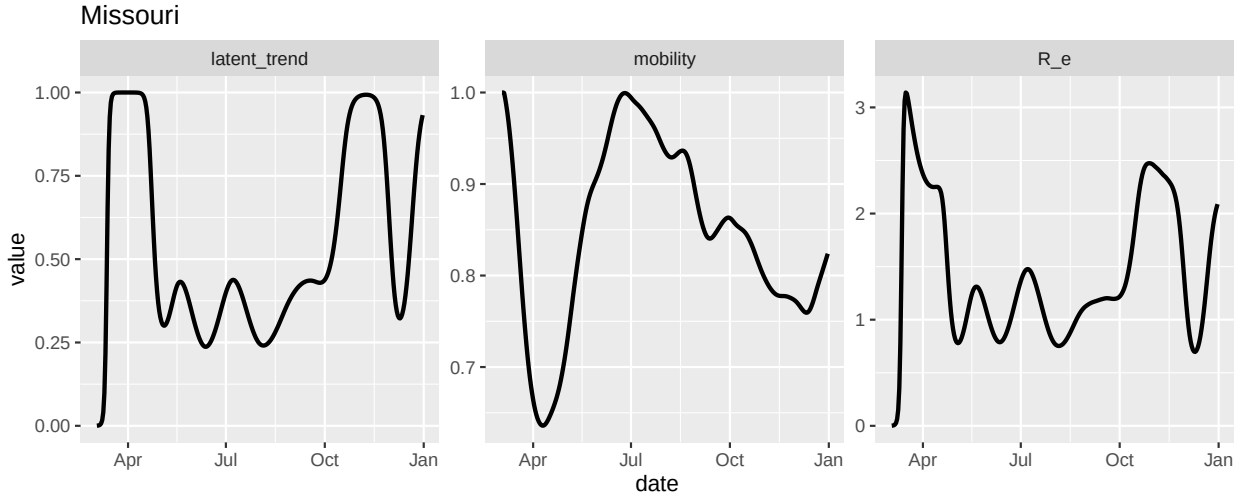

Louisiana
latent_trend
mobility
R_e
value
0.8
0.6
0.4
0.9
0.8
0.7
0.6
3
2
1
Apr
Jul
Oct
Jan
date

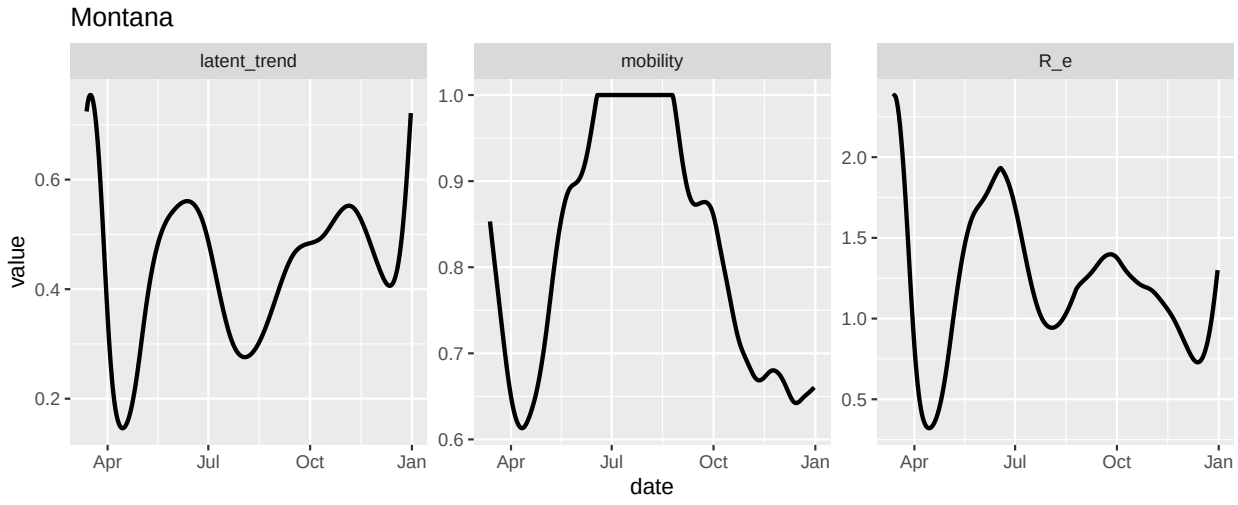

Maine
latent_trend
mobility
R_e
value
1.00
0.75
0.50
0.25
0.9
0.8
0.7
0.6
2.0
1.5
1.0
0.5
Apr
Jul
Oct
Jan
date

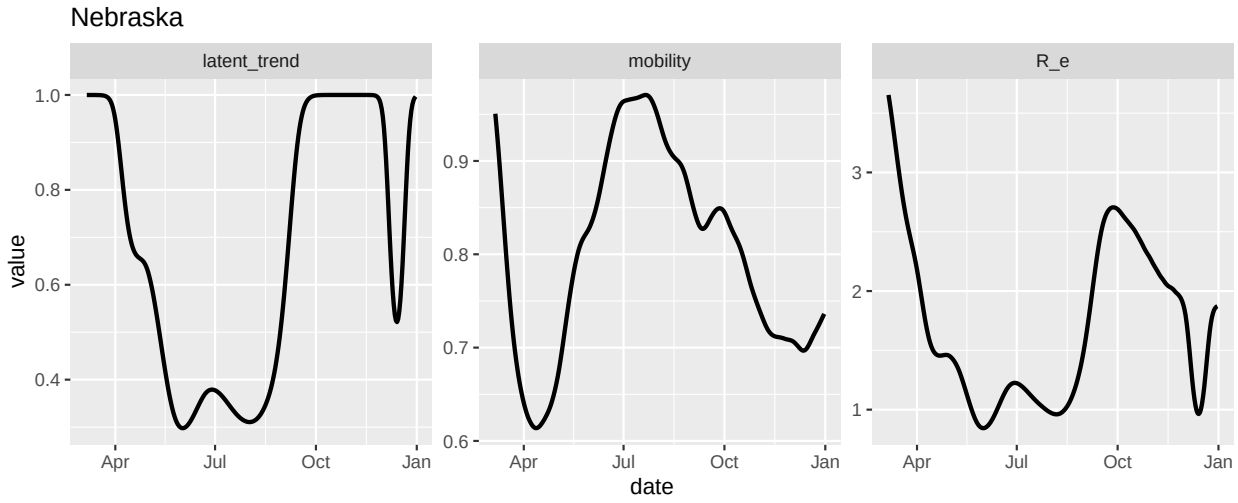

Maryland
latent_trend
mobility
R_e
value
1.00
0.75
0.50
0.25
0.9
0.8
0.7
0.6
0.5
2.5
2.0
1.5
1.0
0.5
Apr
Jul
Oct
Jan
date

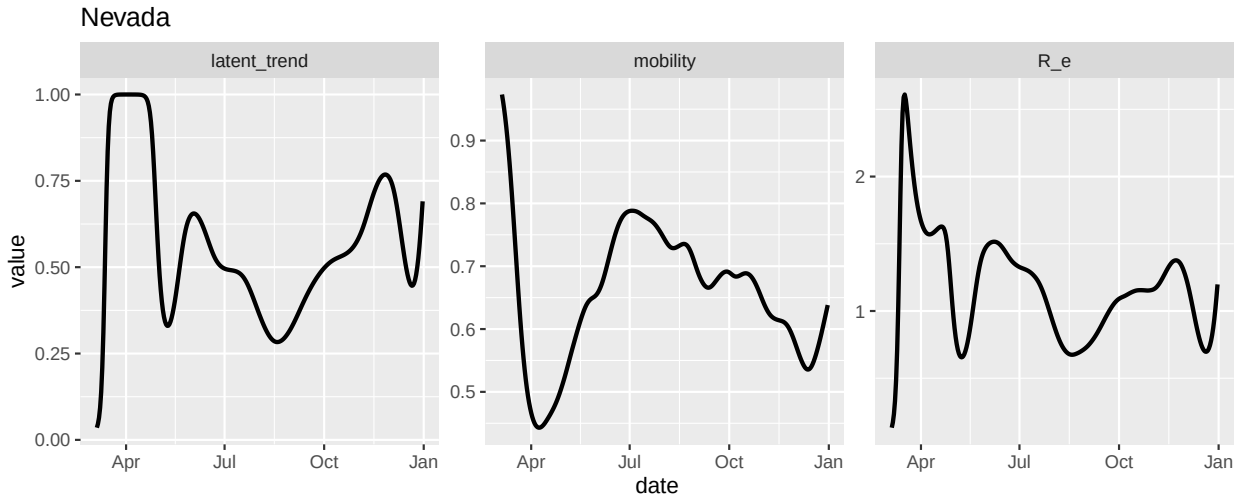

Massachusetts
latent_trend
mobility
R_e
value
date
Apr
Jul
Oct
Jan
Michigan
Minnesota

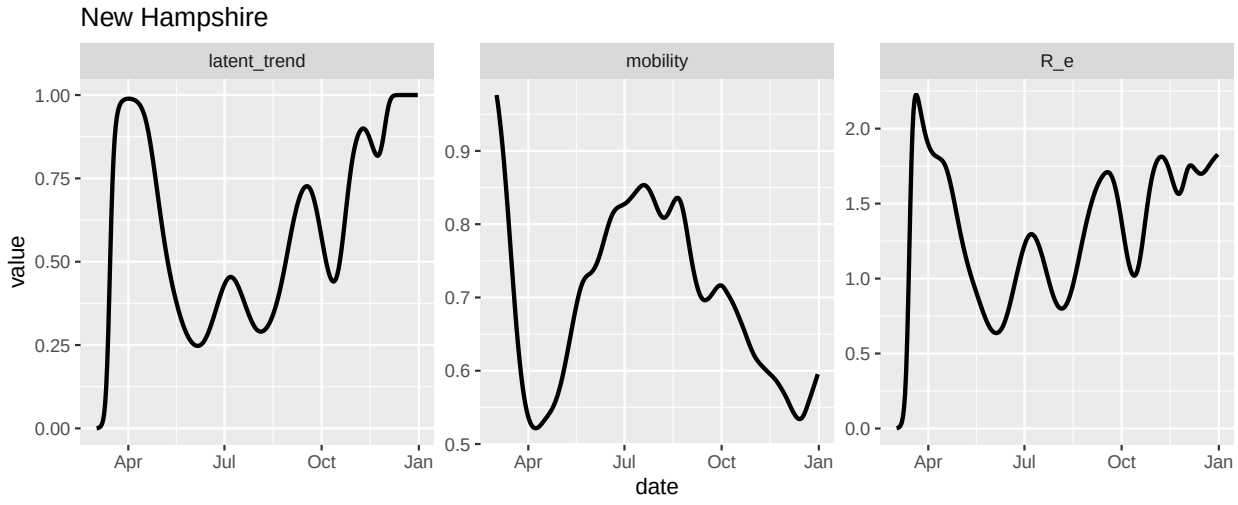

Mississippi
latent_trend
mobility
R_e
value
date
Apr
Jul
Oct
Jan

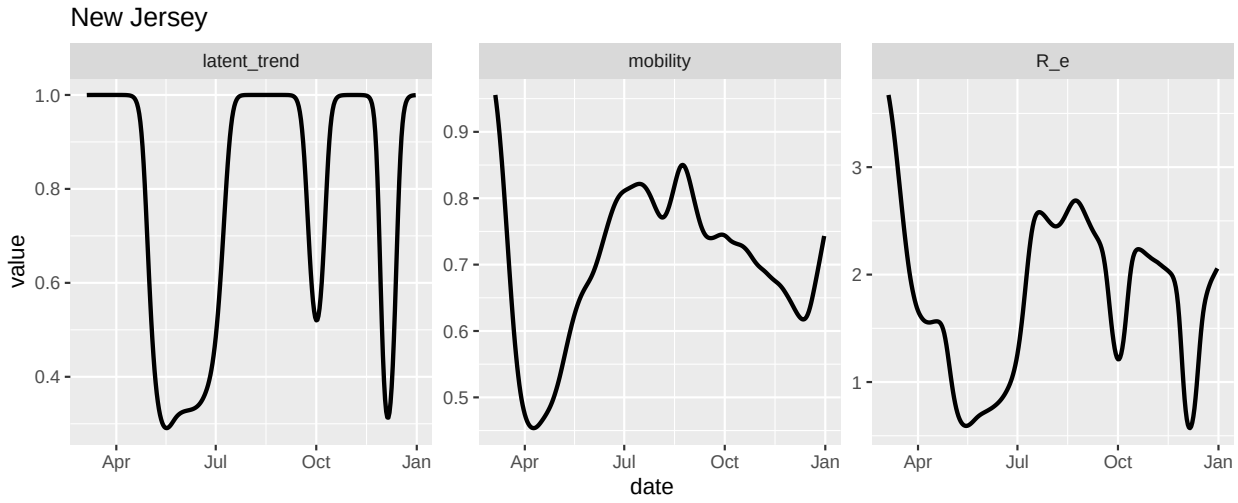

Missouri
latent_trend
mobility
R_e
value
date
Apr
Jul
Oct
Jan

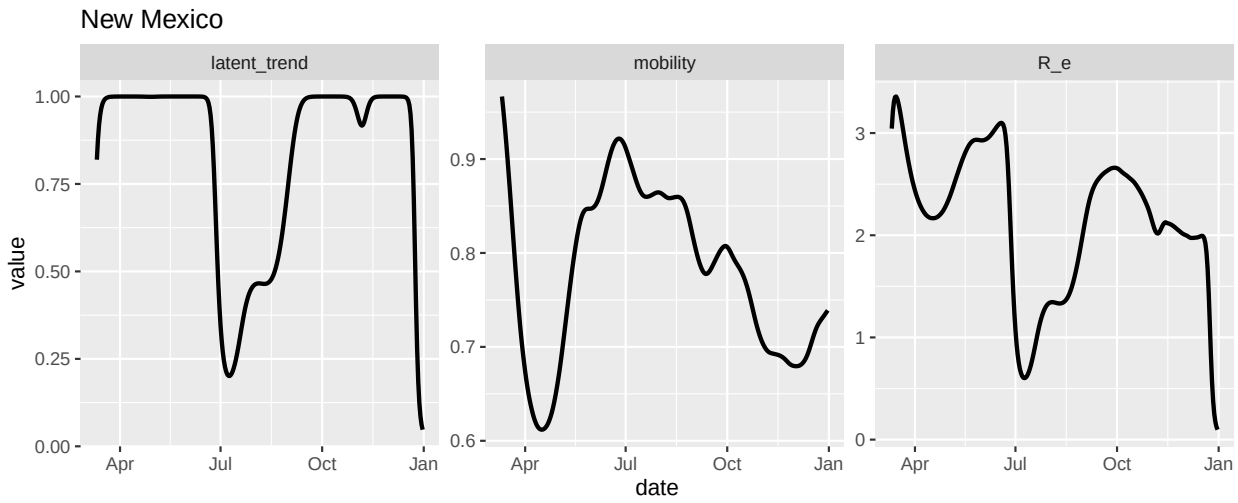

Montana
latent_trend
mobility
R_e
value
date
Apr
Jul
Oct
Jan


Nebraska
latent_trend
mobility
R_e
value
1.0
0.8
0.6
0.4
0.9
0.8
0.7
0.6
3
2
1
Apr
Jul
Oct
Jan
date


Nevada
latent_trend
mobility
R_e
value
1.00
0.75
0.50
0.25
0.00
0.9
0.8
0.7
0.6
0.5
2
1
Apr
Jul
Oct
Jan
date


New Hampshire
latent_trend
mobility
R_e
value
1.00
0.75
0.50
0.25
0.00
0.9
0.8
0.7
0.6
0.5
2.0
1.5
1.0
0.5
0.0
Apr
Jul
Oct
Jan
date


New Jersey
latent_trend
mobility
R_e
value
1.0
0.8
0.6
0.4
0.9
0.7
0.5
3
2
1
Apr
Jul
Oct
Jan
date


New Mexico
latent_trend
mobility
R_e
value
1.00
0.75
0.50
0.25
0.00
0.9
0.8
0.7
0.6
3
2
1
0
Apr
Jul
Oct
Jan
date