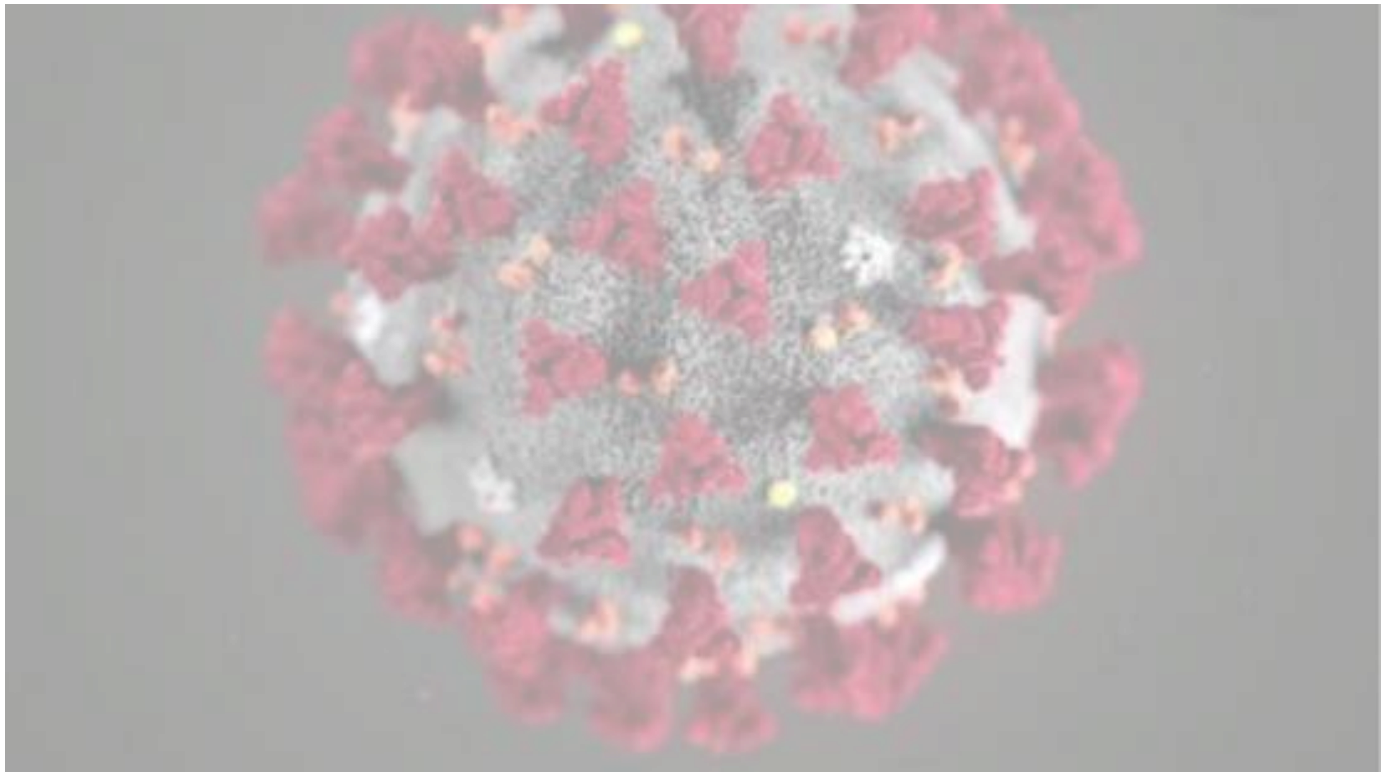
Pandemic Excess Deaths using forecasting methods

Nuria Diaz-Tena

Collaborators:

Davit Sargsyan, Javier Cabrera, Michael Katehakis, Jingyu Cai, Dhammika Amaratunga, Ana Altunya, I-Ming Chiu, Yajie Duan, Chun Pang Lin, Jin Wan & The Rutgers COVID Analytics Research Group



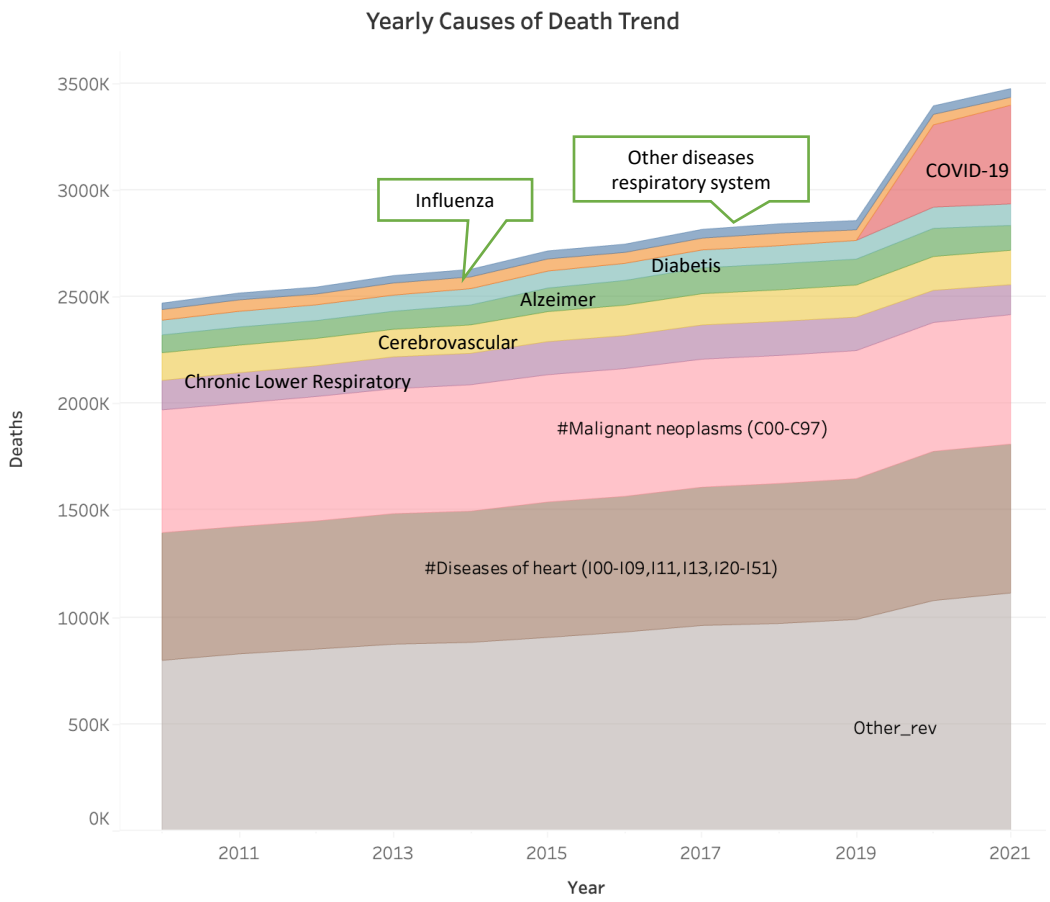
RUTGERS UNIVERSITY

How did the COVID Pandemic affect the US?

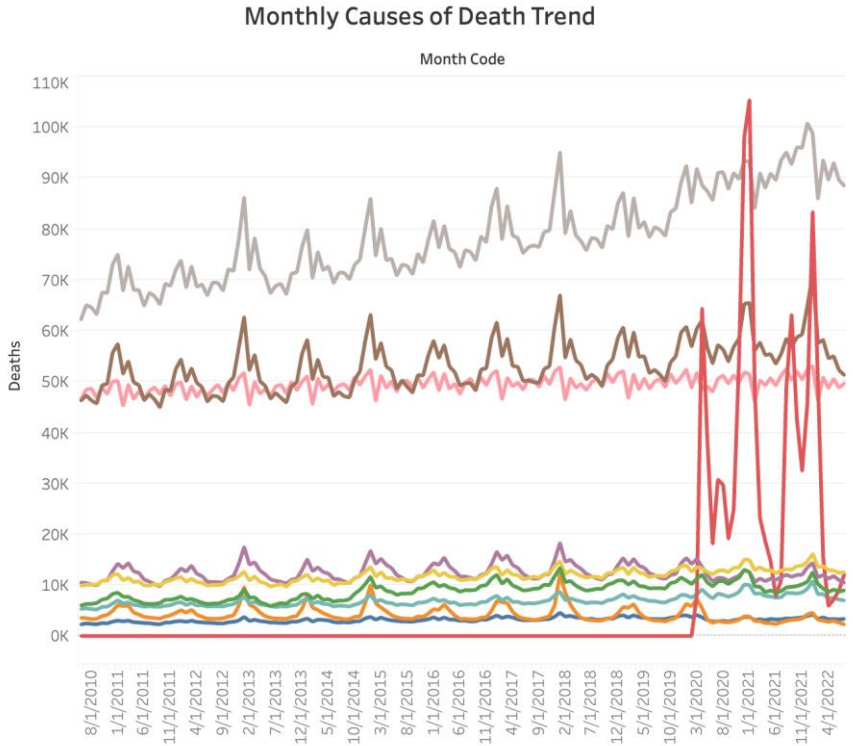
- Background of Deaths in the US
  - By state
  - By cause of death
- Definition of Excess Death
- Forecasting Methods
- Excess deaths compared to COVID deaths over time
- Major causes of death affected by COVID-19
- Adding an additional dimension to the data set for forecasting
- States more affected by COVID-19

Data Sources:
CDC Mortality data for US (from 2010 data July 2022)

Yearly death increased between 1% to 3% before COVID-19. It increased 19% in 2020



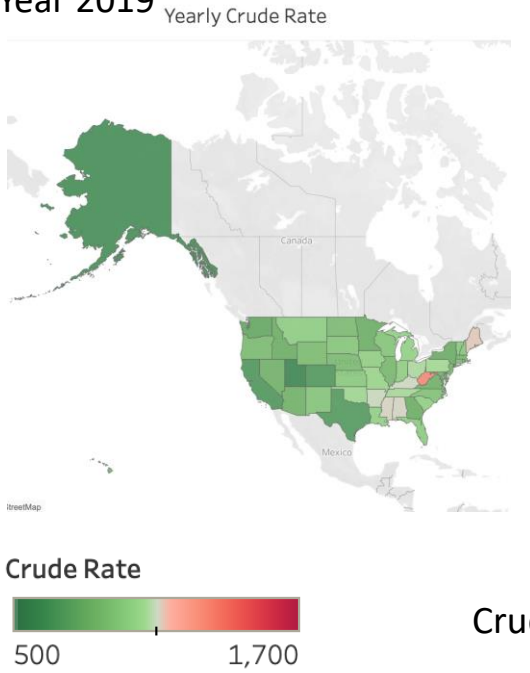
- The major causes of death are diseases of the heart and cancer.
- COVID-19 was the third cause of death in 2020



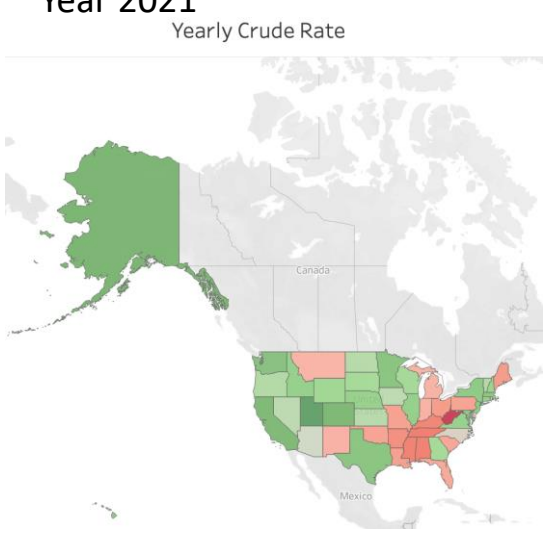
West Virginia, Maine, Mississippi, Alabama, Kentucky, Tennessee and Arkansas are the states with the largest Crude rate in 2019

- West Virginia had a crude rate in 2019 of 136. It increased to 1619 in 2021
- California, Texas and Florida are the states with the largest yearly deaths – because they are the most populous states

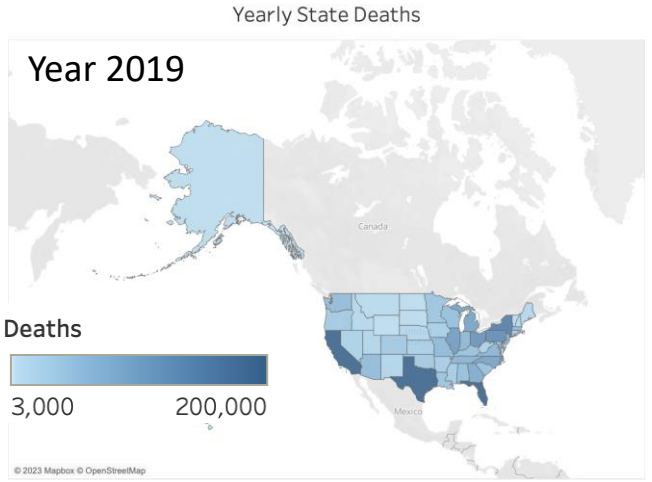
Year 2019

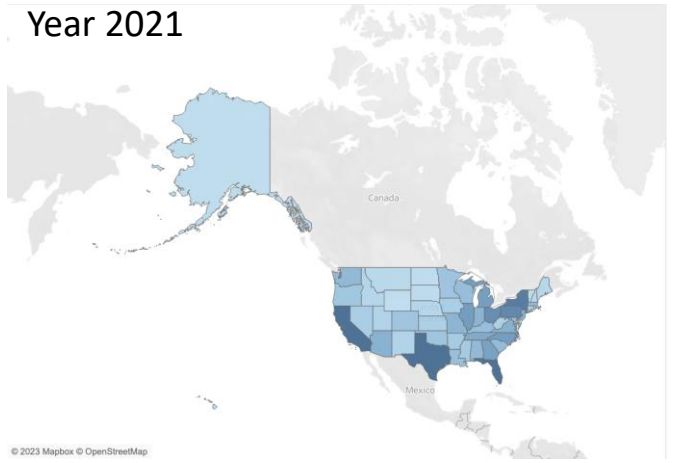


Year 2021

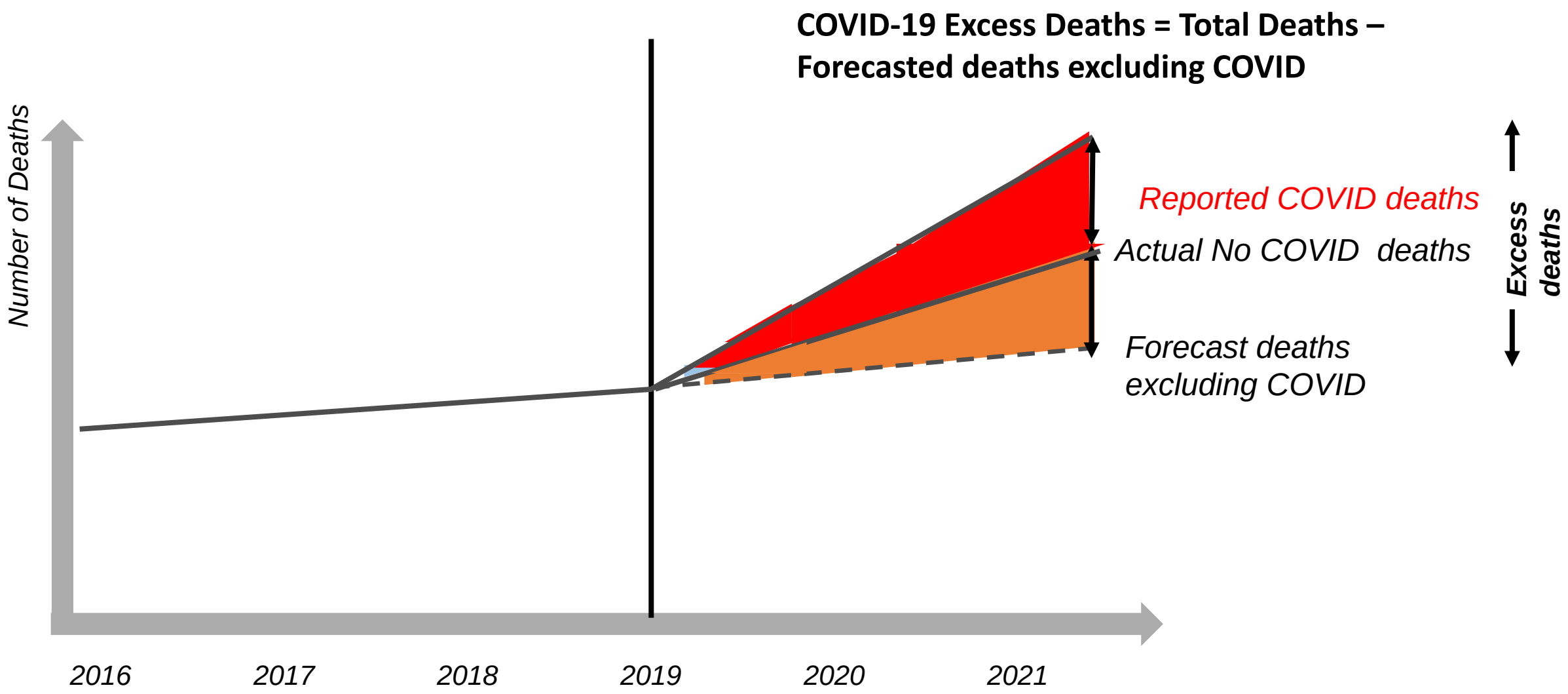


Crude Rate = 100,000* Deaths/Population





What are excess deaths?



Spatial Time Series Forecasting Methods

Spatial Time Series Forecasting

1. **Spatial Forecast Exponential Smoothing**
2. VAR MODEL (Vector Autoregressive Models) A VAR model is a generalization of the univariate autoregressive model for forecasting a vector of time series. It comprises one equation per variable in the system. The right-hand side of each equation includes a constant and lags of all of the variables in the system.
3. VARIMA (Vector Autoregressive Integrated Moving Average) Model
4. GSTAR-SUR Generalized Space Time Autoregressive with Seasonal
5. FFNN-VAR Feedforward Neural Network – Vector Autoregressive
6. FFNN-GSTAR Feedforward Neural Network - Generalized Space-Time Autoregressive
7. Bootstrapping Time series
8. Dynamic harmonic regression (Fourier terms) TBATS model (combination of Exponential smoothing and Fourier terms – more flexible to recent changes due to ES weighting scheme)
9. ….

2-dimensional VAR(1)

$$y_{1,t} = c_1 + \phi_{11,1} y_{1,t-1} + \phi_{12,1} y_{2,t-1} + e_{1,t}$$
$$y_{2,t} = c_2 + \phi_{21,1} y_{1,t-1} + \phi_{22,1} y_{2,t-1} + e_{2,t}$$

Monthly Exponential Smoothing used at the state and cause of death level

Exponential smoothing state space model with Box-Cox transformation, ARMA errors, Trend and Seasonal components

$$\hat{y}_{t+1} = L_t + kT_t + S_{t+k-m}$$

Level Trend Season

$$L_t = \alpha(y_t - S_{t-m}) + (1-\alpha)(L_{t-1} + T_{t-1}),$$

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1},$$

$$S_t = \gamma(y_t - L_t) + (1-\gamma)S_{t-m}.$$

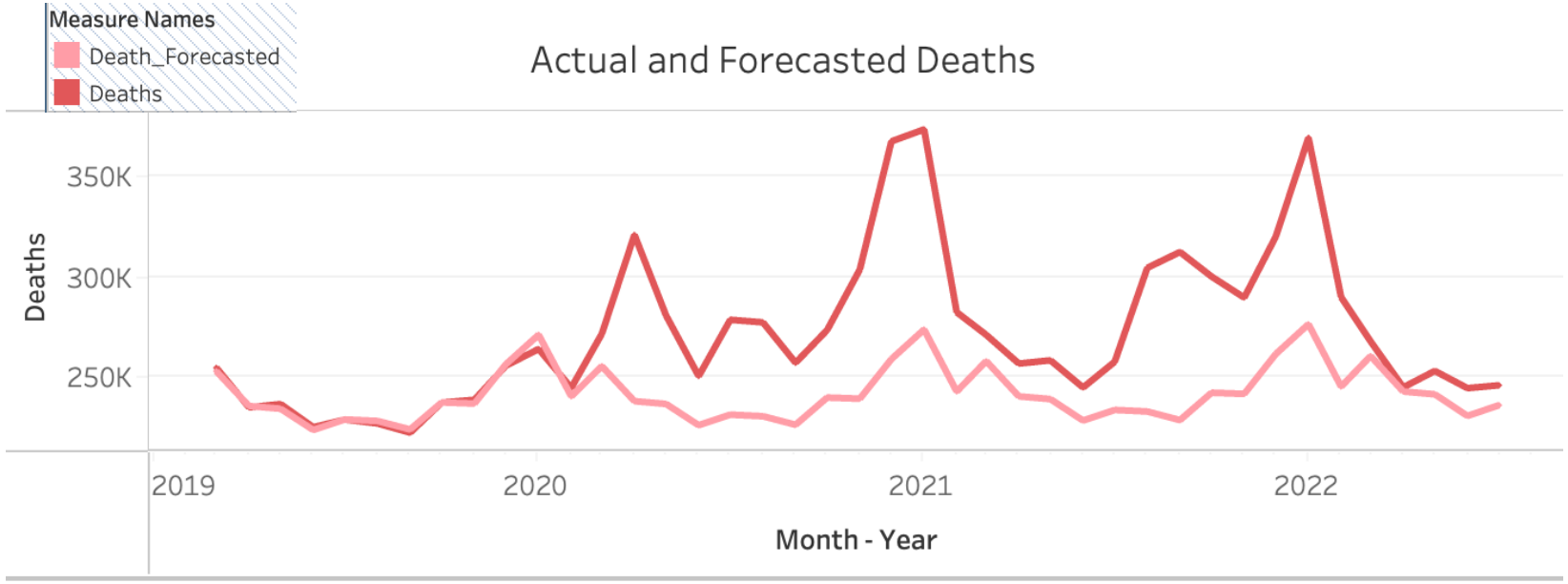
where α, β, and γ are the three smoothing parameters

t – time in months

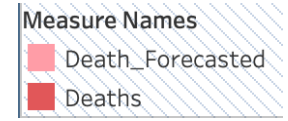
Forecasts for each state and cause of death

Overall Forecast

Monthly Historical data from 2010 till 2019 February was used to train the model. The forecast starts in 2019 March. The actual deaths and the forecasted are very close during the first year of the forecast (prior to COVID-19). The Actual is larger than the forecasted starting in 2020 March

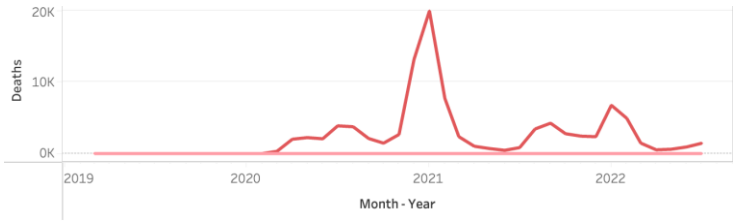


California Forecast

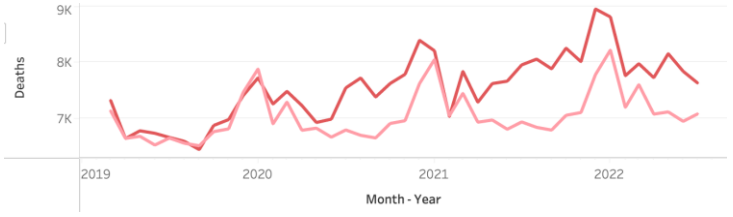


- ✓ The Cancer forecast is overestimated
- ✓ The other diseases are on target on the year prior to COVID.
- ✓ The forecast after COVID-19 is larger than the actual (Reduced Deaths) or smaller than the actual (Excess Deaths)

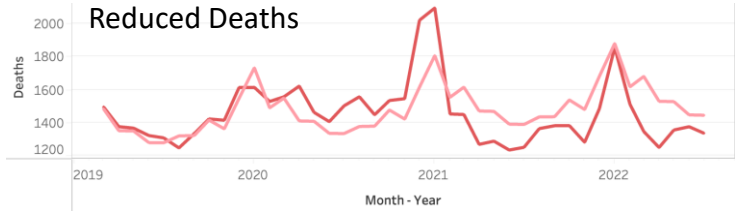
COVID-19 – Excess Deaths



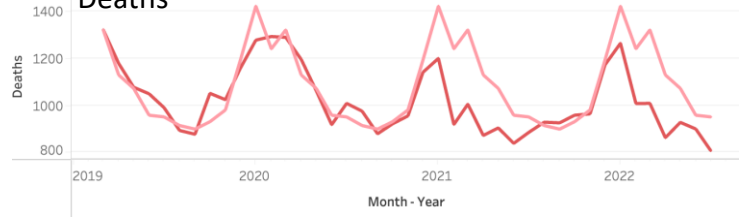
Other – Excess Deaths



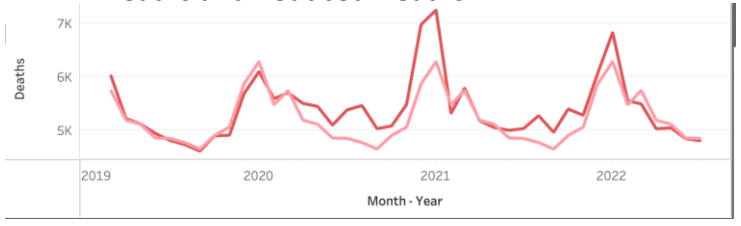
Alzheimer's – Mixture of Excess Deaths and Reduced Deaths



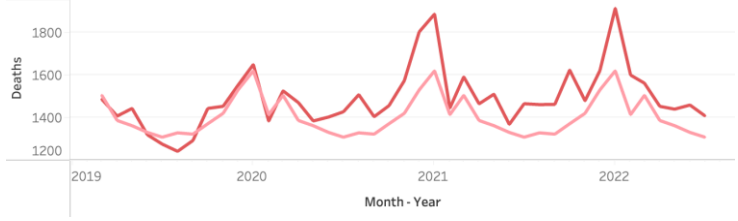
Chronic Lower Respiratory – Reduced Deaths



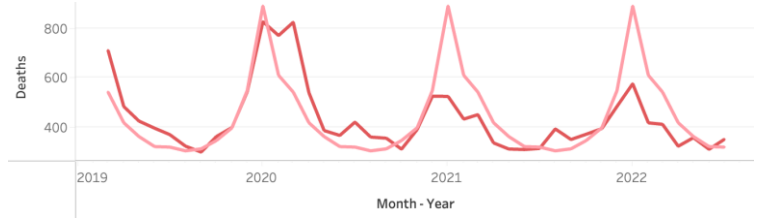
Diseases of the Heart – Mixture of Excess Deaths and Reduced Deaths



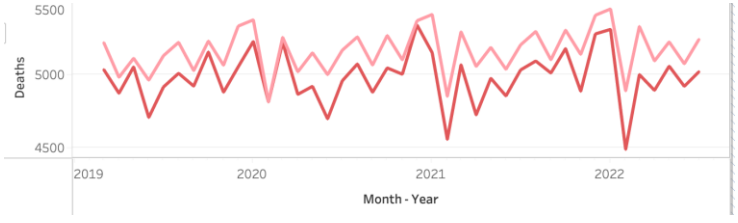
Cerebrovascular – Excess Deaths



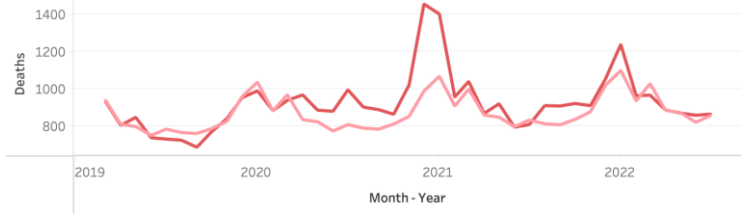
Influenza and Pneumonia – Reduced Deaths



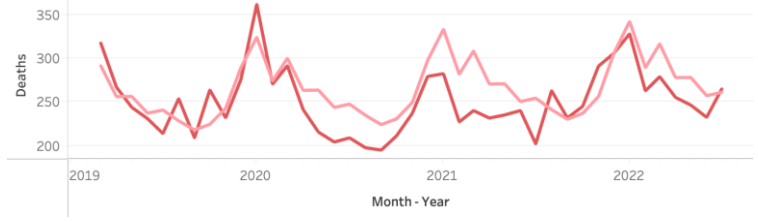
Cancer – Overestimated



Diabetes – Excess Deaths



Other diseases Respiratory system – Reduced Deaths



Misclassified COVID deaths have decreased over time

- New protocols and tests had to be introduced as a consequence of COVID-19 -a new disease that become the 3rd cause in 2020 in the US. We were expecting misclassification cases
- The misclassified cases dropped over time

Data at Year, Month, Cause of death level Prediction model

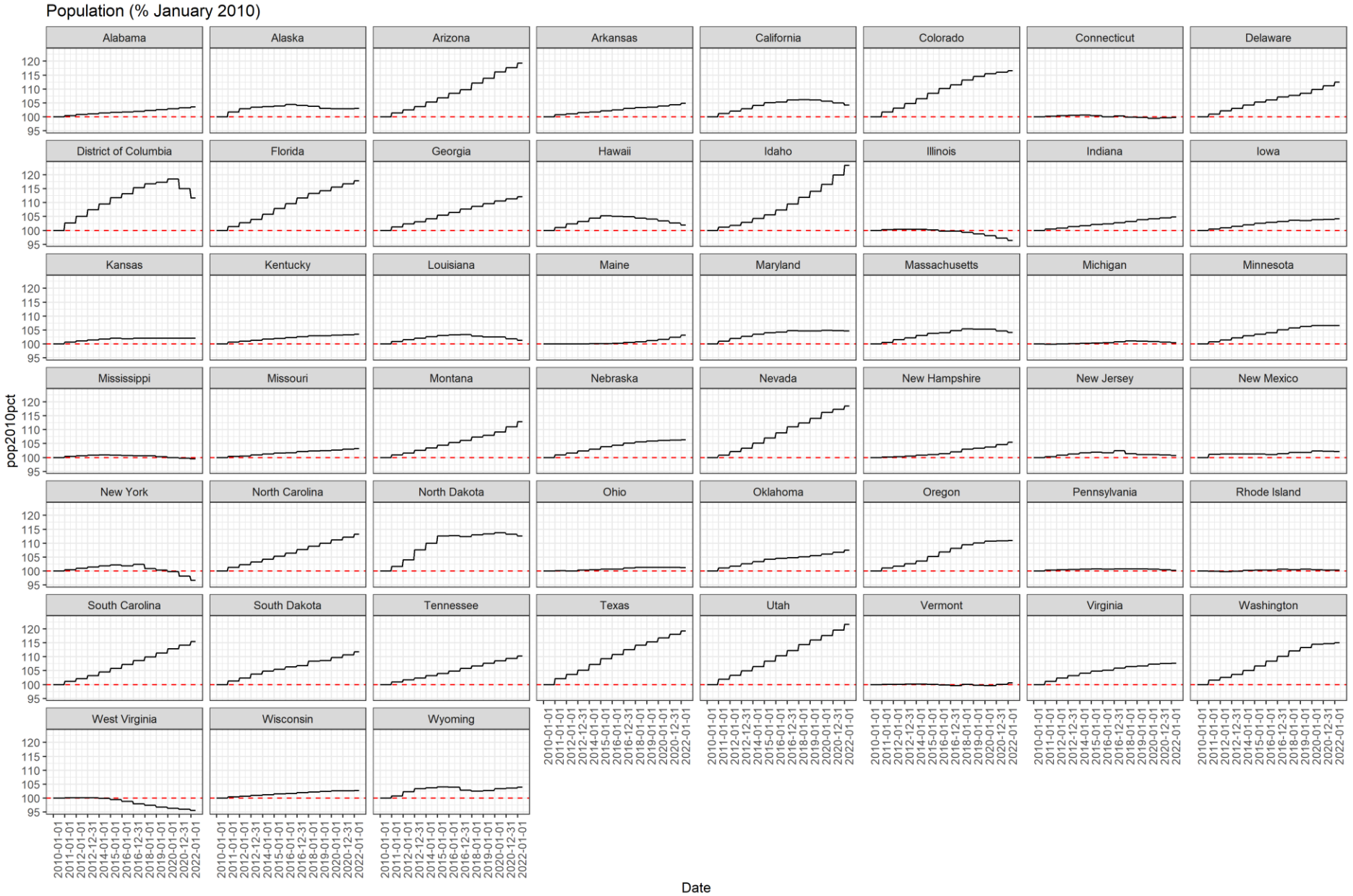
Data collected till July 2022

Year	COVID Deaths	Excess Deaths	Pct unaccounted
2020	385,072	497,157	29%
2021	462,197	546,012	18%
2022 till July	182,195	181,581	0%
Till July 2022	**1,030,264**	**1,224,750**	**19%**

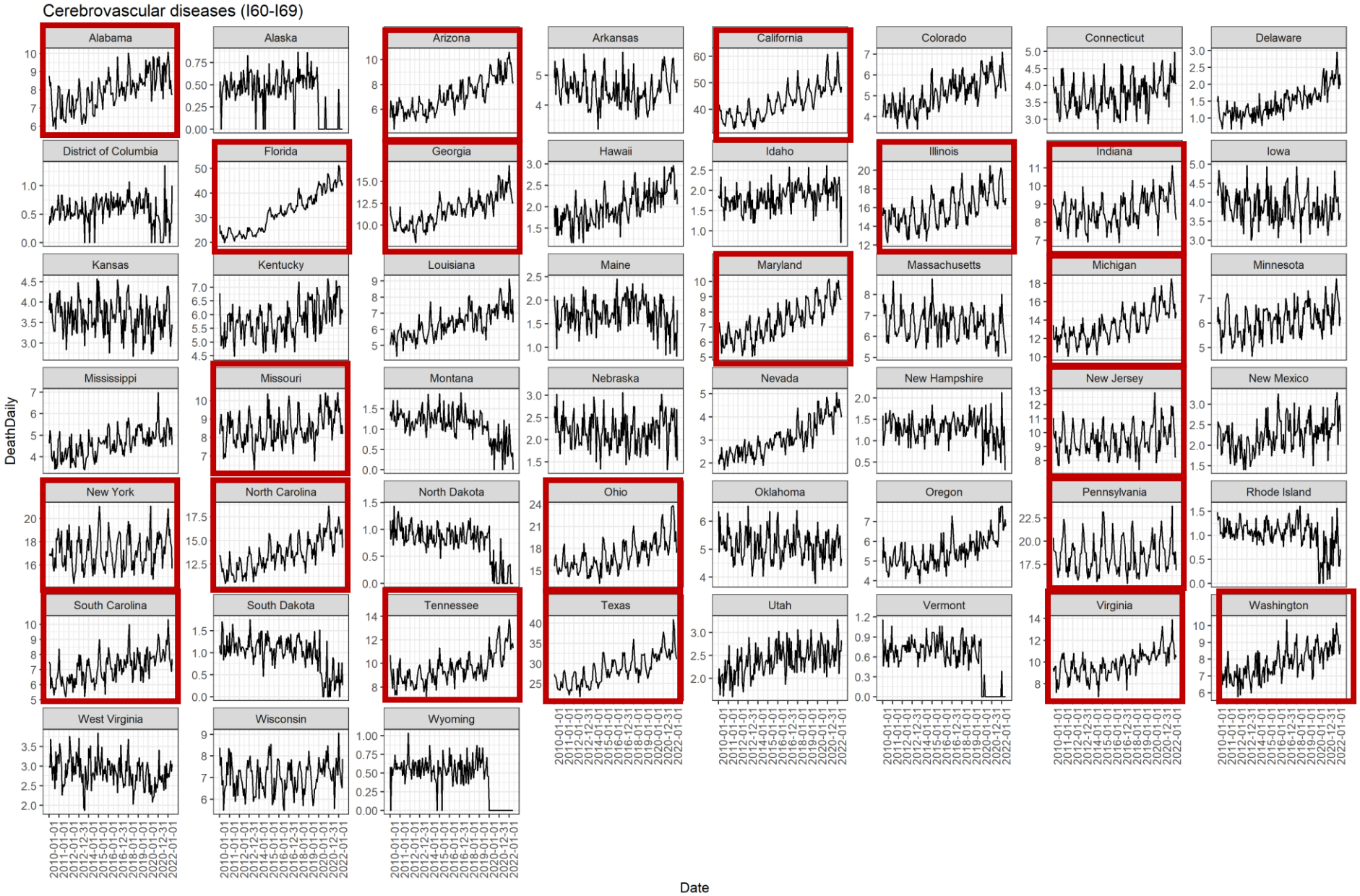
Data processing and subsetting

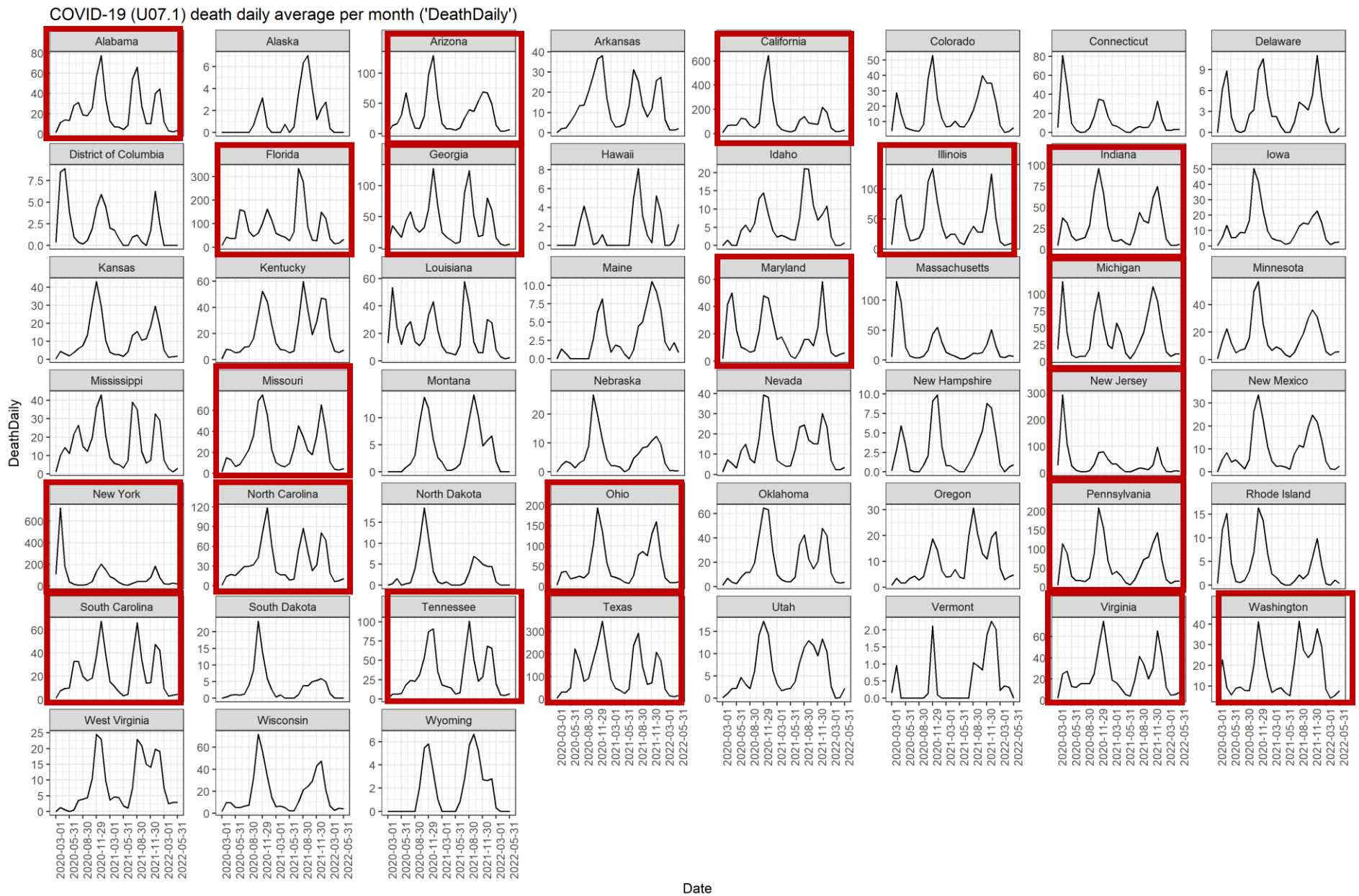
- DeathDaily = Death/daymonth (i.e., average daily death = average monthly death/number of days in that month)
- Removed records: Date >= 2022-07-01
- Removed states with max(DeathDaily) <10 in Cerebrovascular diseases (I60-I69). 20 out of 51 states left.
- Divided data into training (before 2019/03/01), testing (2019/03/01 to 2020/02/28 or 29) and Covid-19 (2020/03/01 to 2022/06/01)
- Calculated Rate (no more zeros) = 10^5*(Deaths/(daymonth*Population)
  - NOTE: if there were zeros then Rate = 10^5*(Deaths + 1)/(daymonth*(Population + 1))

Population dynamics (% Jan 2010) in the US 51 states between Jan 2010 and June 2022



Only the states that reached at least 10 CVD deaths daily were kept in the analysis




COVID-19 (U07.1) death daily average per month ('DeathDaily')
DeathDaily
Date
Alabama
Alaska
Arizona
Arkansas
California
Colorado
Connecticut
Delaware
District of Columbia
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
New York
North Carolina
North Dakota
Ohio
Oklahoma
Oregon
Pennsylvania
Rhode Island
South Carolina
South Dakota
Tennessee
Texas
Utah
Vermont
Virginia
Washington
West Virginia
Wisconsin
Wyoming

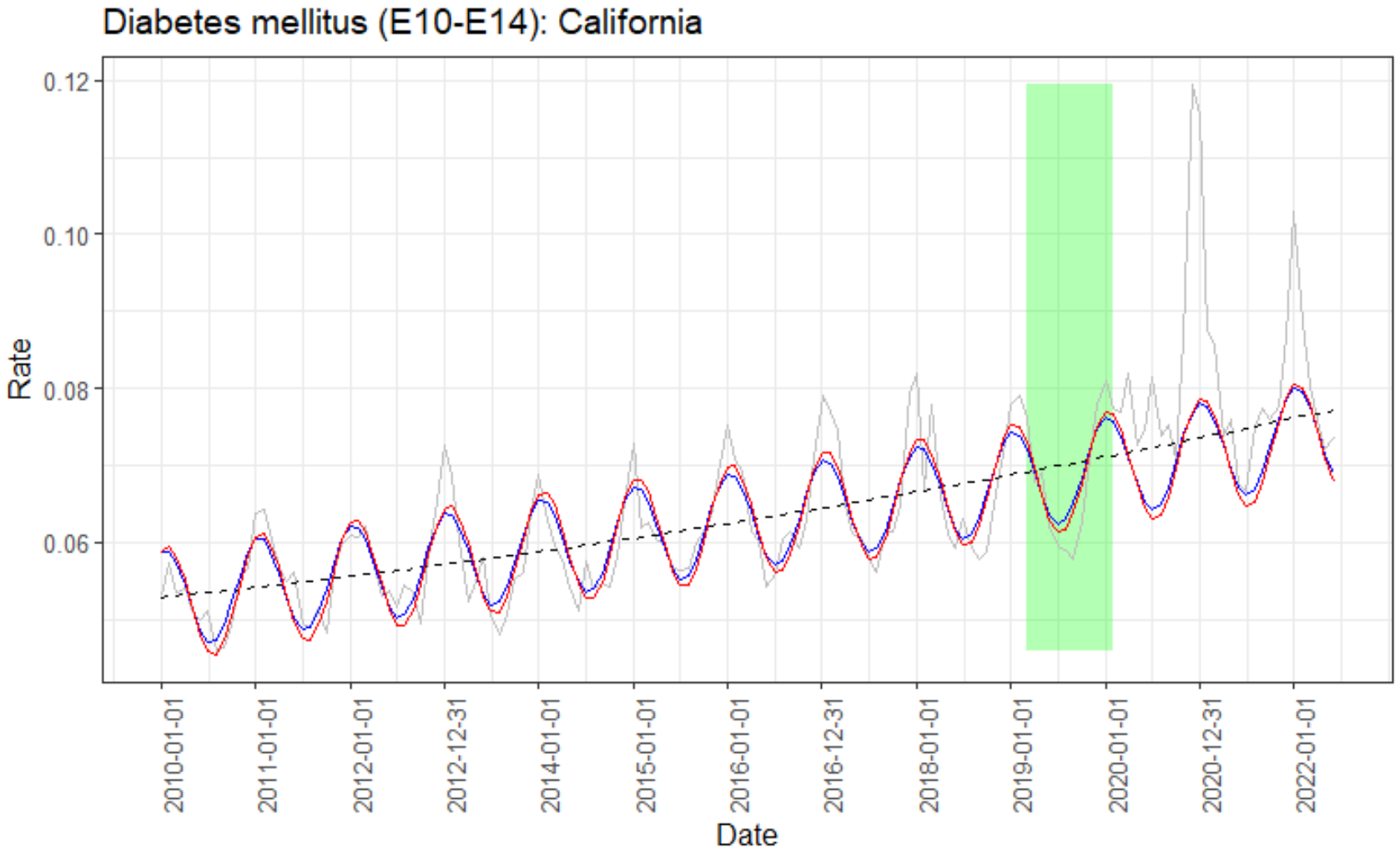
Nonlinear Mixed-Effects Model

- Rate ~ b0 + b1*dt + b2*dt^2 + b3*sin(b4 + b5*dt) | State

  where b0: y-intersept; b1: linear trend; b2: quadratic term; b3: wave magnitude; b4: wave offset left; b5: wavelength.

Cause of death	b0	b1	b2	b3	b4	b5
Alzheimer disease (G30)	0.4277	-5.40E-05	2.04E-09	0.0119	15.14	0.0171
Cerebrovascular diseases (I60-I69)	0.4727	-4.78E-05	1.61E-09	0.0107	14.67	0.0172
Chronic lower respiratory diseases (J40-J47)	-0.1350	2.89E-05	-7.55E-10	0.0226	13.80	0.0172
Diabetes mellitus (E10-E14)	0.0154	3.97E-06	-4.42E-11	0.0070	13.43	0.0172
Diseases of heart (I00-I09,I11,I13,I20-I51)	1.1233	-7.92E-05	2.72E-09	0.0547	14.26	0.0172
Influenza and pneumonia (J09-J18)	-0.3035	4.26E-05	-1.29E-09	0.0170	14.89	0.0171
Malignant neoplasms (C00-C97)	0.1835	4.27E-05	-1.34E-09	0.0096	14.47	0.0172
Other_rev	0.1453	3.95E-05	-2.92E-13	0.0501	14.64	0.0172
Other diseases of respiratory system (J00-J06,J30- J39,J67,J70-J98)	0.0644	-7.29E-06	3.33E-10	0.0043	14.42	0.0172

Model Fits Example: Diabetes in CA



- Grey curve: observed daily death rate
- **Black** dashed curve: LM prediction (with quadratic term)
- Blue curve: NLME prediction using all states
- Red curve: NLS prediction using only CA data

Nonlinear Mixed-Effects Model
Mean Absolute Percent Error

MAPE = 100*mean(abs(Rate - prd3)/(Rate)))

Cause of Death	Training	Testing	COVID-19
Alzheimer disease (G30)	8.65	14.21	28.05
COVID-19 (U07.1)	6.64	7.38	14.91
Cerebrovascular diseases (I60-I69)	5.06	5.39	5.99
Chronic lower respiratory diseases (J40-J47)	6.11	5.85	14.80
Diabetes mellitus (E10-E14)	3.11	2.93	6.20
Diseases of heart (I00-I09,I11,I13,I20-I51)	15.66	14.61	39.57
Influenza and pneumonia (J09-J18)	2.16	2.12	3.01
Malignant neoplasms (C00-C97)	2.80	2.43	7.66
Other diseases of respiratory system (J00-J06,J30- J39,J67,J70-J98)	8.65	14.21	28.05
Other_rev	9.39	10.91	Inf

Nonlinear Mixed-Effects Model
Predicted Death

Deaths_prd := prd3*daysmonth*Population*10^(-5)

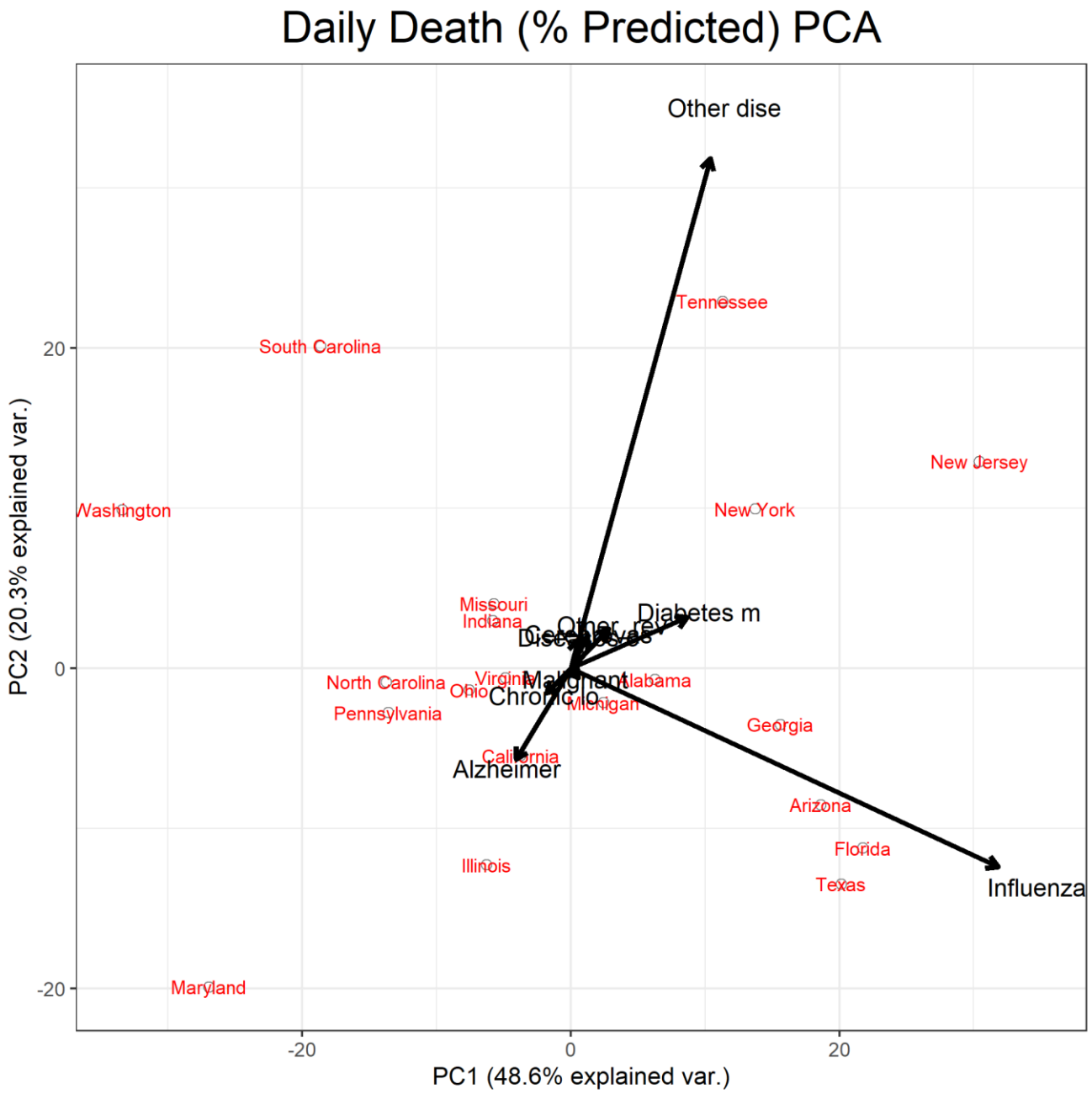
Excess Death (% predicted) = 100*sum(Deaths - Deaths_prd)/sum(Deaths_prd)

Cause of Death	Training	Testing	COVID-19
Alzheimer disease (G30)	8.65	14.21	28.05
COVID-19 (U07.1)	6.64	7.38	14.91
Cerebrovascular diseases (I60-I69)	5.06	5.39	5.99
Chronic lower respiratory diseases (J40-J47)	6.11	5.85	14.80
Diabetes mellitus (E10-E14)	3.11	2.93	6.20
Diseases of heart (I00-I09,I11,I13,I20-I51)	15.66	14.61	39.57
Influenza and pneumonia (J09-J18)	2.16	2.12	3.01
Malignant neoplasms (C00-C97)	2.80	2.43	7.66
Other diseases of respiratory system (J00-J06,J30- J39,J67,J70-J98)	8.65	14.21	28.05
Other_rev	9.39	10.91	Inf

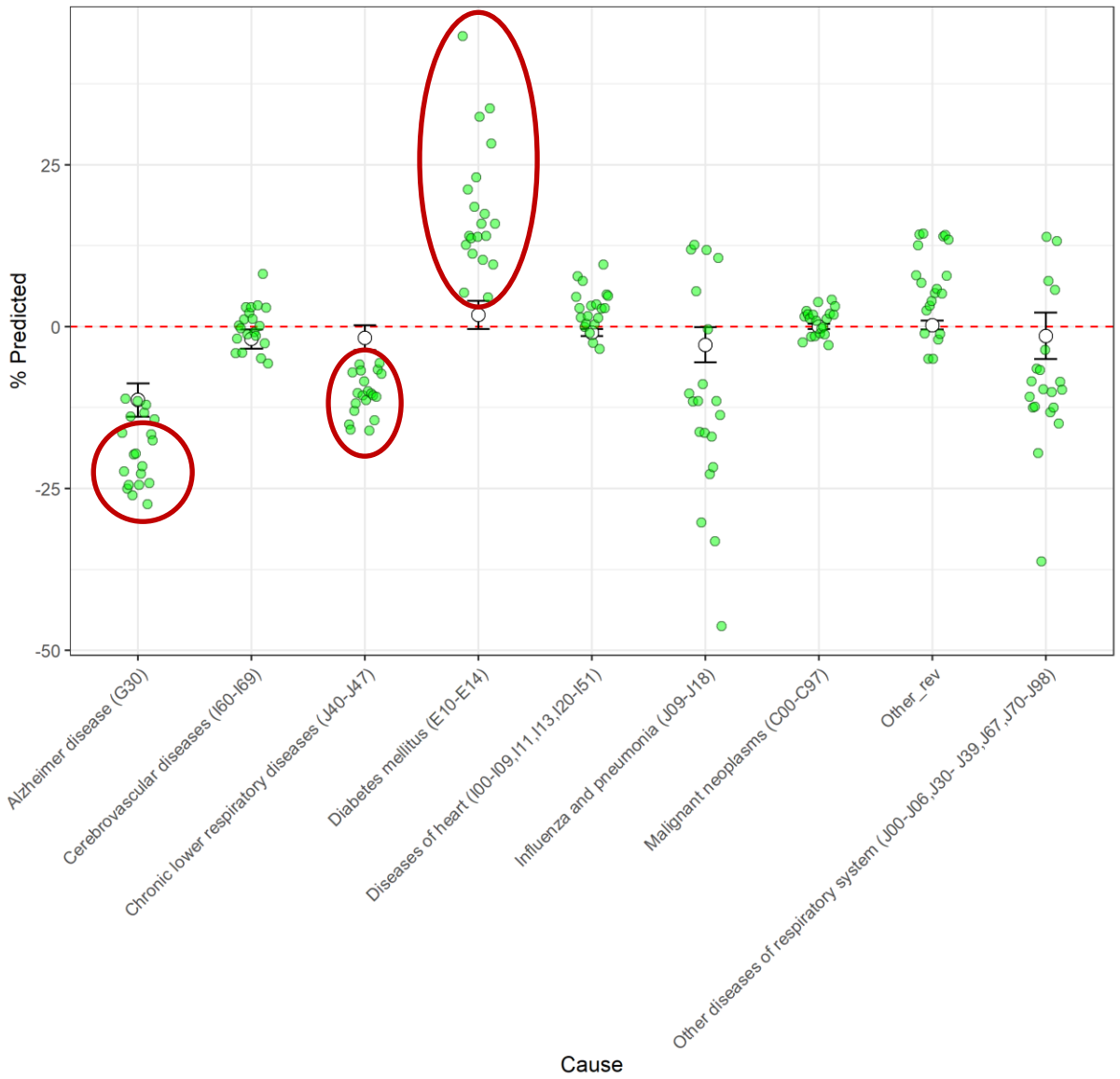
State	Observed Death During the Testing Months (2019/03/01 to 2020/02/28, % Predicted)								
	Alzheimer disease (G30)	Cerebrovascular diseases (I60-I69)	Chronic lower respiratory diseases (J40-J47)	Diabetes mellitus (E10-E14)	Diseases of heart (I00-I09,I11,I13,I20-I51)	Influenza and pneumonia (J09-J18)	Malignant neoplasms (C00-C97)	Other_rev	Other diseases of respiratory system (J00-J06,J30-J39,J67,J70-J98)
Alabama	-11.1	1.0	-4.3	7.7	-1.6	-1.7	-2.3	1.1	-11.8
Arizona	-9.8	-4.6	-10.0	-2.5	-0.3	11.9	0.9	0.3	-7.8
California	-6.3	-2.0	-0.7	-0.8	-1.1	-2.0	0.2	2.1	-0.1
Florida	-19.3	-0.4	-3.6	-0.2	-1.2	3.2	0.3	0.2	0.0
Georgia	-16.9	-6.6	-6.8	-2.4	1.5	-1.7	0.8	3.0	-9.8
Illinois	-11.1	-0.1	-1.6	-4.1	-1.4	-3.2	-0.5	-1.2	-12.3
Indiana	-13.2	5.2	2.1	-6.3	-0.6	-3.6	1.0	-1.0	-1.1
Maryland	-20.7	1.0	3.7	2.8	-0.1	-10.6	-0.8	-2.5	-7.6
Michigan	-9.8	-0.1	2.4	5.7	-2.6	2.4	0.6	-0.2	2.8
Missouri	3.3	-3.8	-5.5	1.5	-1.2	-12.2	-0.7	-1.9	6.6
New Jersey	-14.0	0.7	-3.1	7.4	-2.5	0.8	-0.5	-0.2	5.0
New York	-7.4	-3.5	2.6	5.5	-1.9	0.6	0.0	1.3	11.0
North Carolina	-8.5	-2.9	-3.8	-4.2	-0.5	-7.5	0.0	-0.1	-8.6
Ohio	-8.8	-0.5	-0.9	1.4	-0.1	-10.0	-0.9	-2.0	-9.1
Pennsylvania	-9.6	-0.4	-1.9	-2.1	-2.7	-6.2	0.5	-1.7	-3.3
South Carolina	-22.4	-8.6	-7.1	2.4	-2.6	-6.7	-0.8	0.5	13.4
Tennessee	-14.9	-2.7	-3.2	10.4	1.5	-8.2	-0.5	2.1	12.9
Texas	-11.6	-7.4	0.3	11.3	-2.2	5.6	2.2	2.2	-10.7
Virginia	-5.8	1.5	9.1	4.0	0.6	-9.3	0.5	0.7	1.5
Washington	-9.3	-4.5	-2.8	-1.3	0.4	2.0	0.0	1.9	0.2

State	Observed Death During COVID-19 (2020/03/01 to 2022/06/01, % Predicted)								
	Alzheimer disease (G30)	Cerebrovascular diseases (I60-I69)	Chronic lower respiratory diseases (J40-J47)	Diabetes mellitus (E10-E14)	Diseases of heart (I00-I09,I11,I13,I20-I51)	Influenza and pneumonia (J09-J18)	Malignant neoplasms (C00-C97)	Other_rev	Other diseases of respiratory system (J00-J06,J30-J39,J67,J70-J98)
Alabama	-16.4	-4.1	-15.1	44.9	4.6	-10.3	-2.4	7.9	-10.8
Arizona	-22.3	-1.8	-15.9	5.3	7.8	11.9	1.6	12.6	-8.5
California	-11.1	0.2	-7.0	12.7	2.9	-11.6	2.5	14.3	-12.5
Florida	-25.0	-0.2	-13.0	21.2	1.4	12.6	2.0	6.8	-12.4
Georgia	-24.4	-4.0	-11.8	14.1	7.1	5.5	1.3	14.4	-6.5
Illinois	-13.9	1.2	-10.2	13.7	0.0	-11.5	-1.5	-1.0	-19.5
Indiana	-26.0	3.0	-5.8	11.3	0.5	-16.3	1.9	2.5	-6.7
Maryland	-19.7	-1.2	-6.8	18.5	1.7	-30.2	-1.4	-5.0	-36.2
Michigan	-19.6	2.2	-10.6	23.1	-0.9	-8.9	1.0	3.2	-9.6
Missouri	-11.5	3.1	-8.4	13.9	3.2	-16.4	3.8	3.9	-3.6
New Jersey	-24.4	1.2	-11.3	32.4	-2.5	11.8	-1.0	-4.9	13.9
New York	-22.7	-0.8	-9.9	15.9	0.4	-0.4	-0.2	5.2	7.1
North Carolina	-21.6	-1.4	-16.0	10.3	3.5	-22.7	0.0	5.8	-13.2
Ohio	-13.3	3.3	-10.3	17.5	1.3	-17.0	-1.2	-2.0	-10.1
Pennsylvania	-12.0	0.1	-10.6	14.1	-3.4	-21.7	1.2	-1.1	-12.4
South Carolina	-27.4	-4.9	-14.4	4.5	2.8	-33.1	-2.8	5.1	5.7
Tennessee	-24.2	8.2	-10.8	33.7	9.6	-11.5	2.0	14.0	13.3
Texas	-16.6	-2.6	-6.6	28.3	2.8	10.7	4.2	14.1	-15.0
Virginia	-17.6	3.0	-5.6	9.6	5.0	-13.6	1.9	7.9	-8.5
Washington	-14.3	-5.7	-7.3	15.9	4.8	-46.2	3.2	13.5	-9.7

- PCA on observed daily death (as % of predicted)
- Influenza death were higher than predicted in TX, FL, AZ, GE and NJ
- Diabetes deaths exceeded the predicted the most in NJ, NY and TN

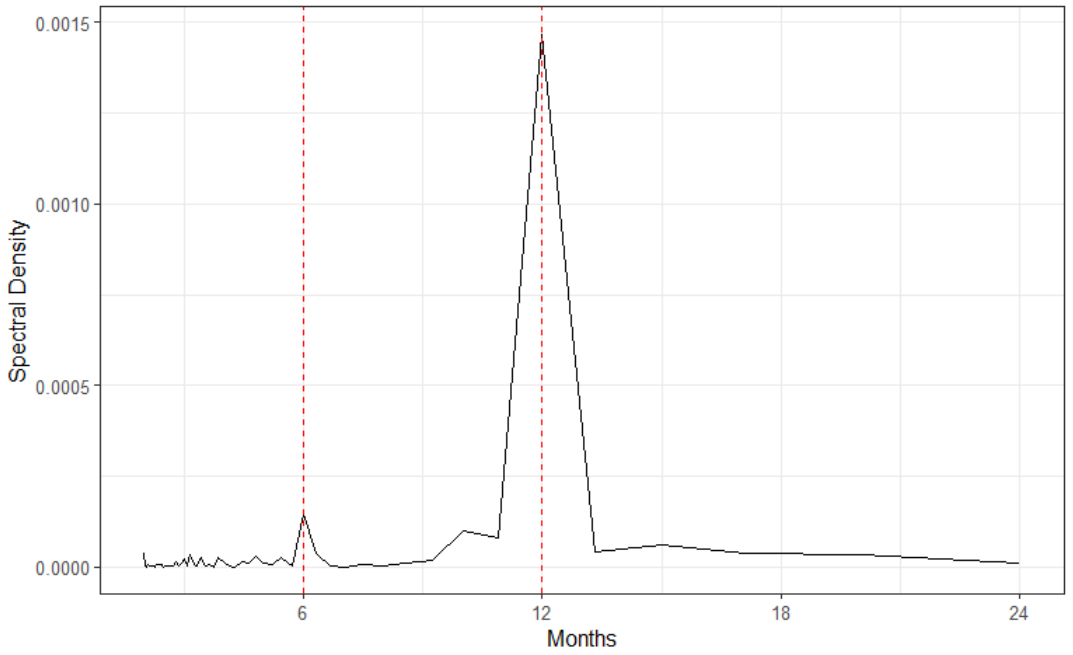


Population dynamics (% Jan 2010) in the US 51 states between Jan 2010 and June 2022

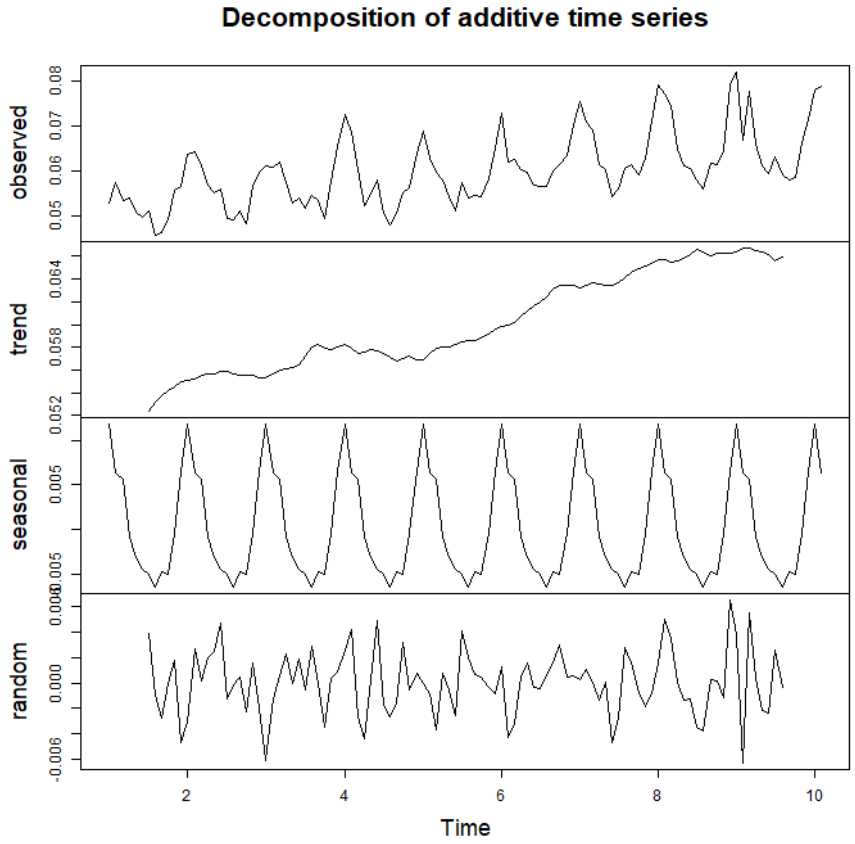


- Error bars indicate 95% CI calculated on the Testing data
- Green dots are the observed deaths as % of predicated, by State
- The model underestimated deaths from Alzheimer even during the Testing years
- During COVID-19 years, the predicted death caused by **Alzheimer, Chronic Lower Respiratory and Influenza** were much lower that expected from the Testing set, while **Diabetes** death was much higher than expected

Classical Seasonal Decomposition by Moving Averages (Diabetes in CA Example)

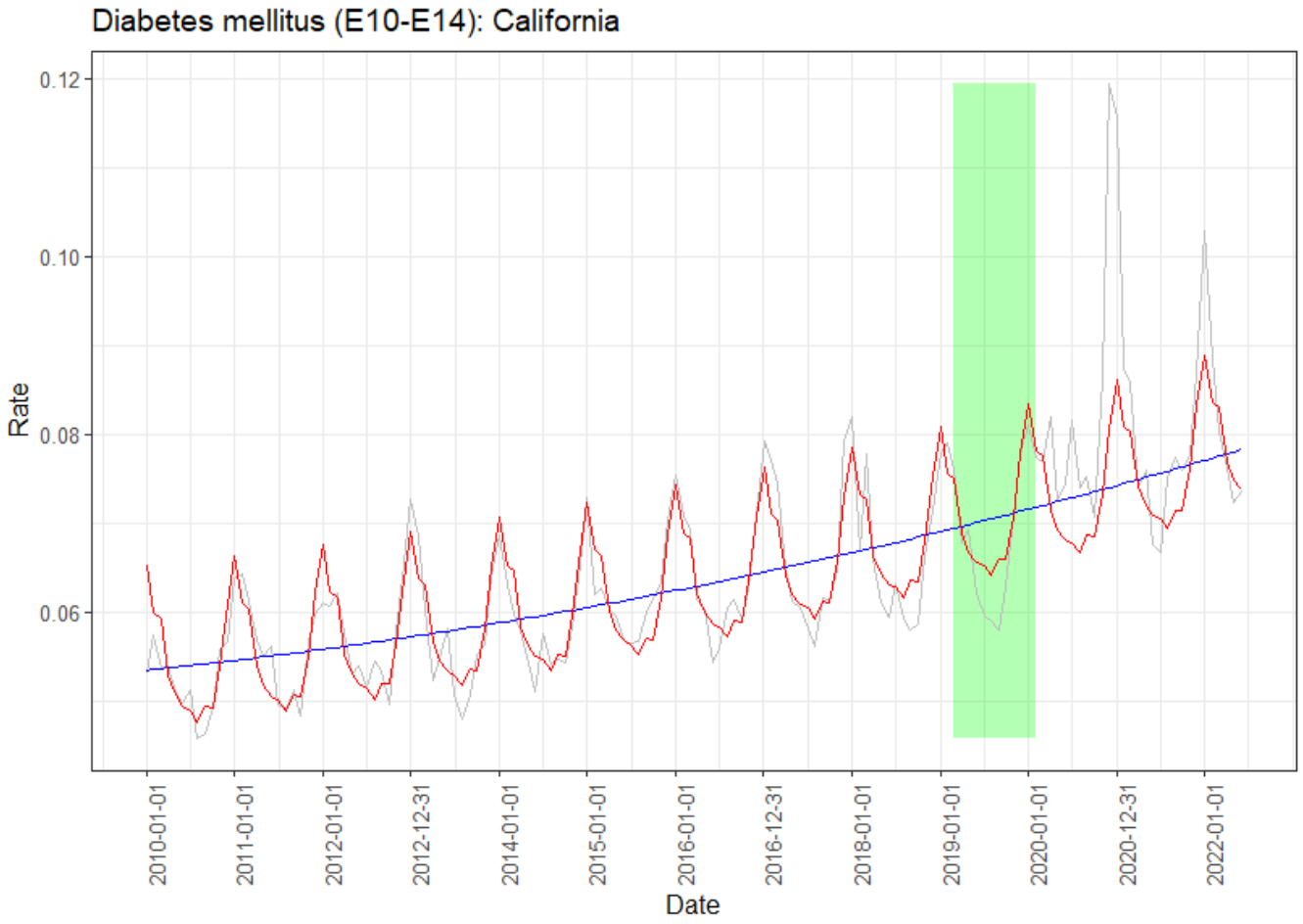


- Spectral density analysis of the time series confirmed periodicity of 12 months (R::spectrum)



- Increasing trend and seasonality captured (R::decompose), with Random component being an order of magnitude smaller than the Seasonal component

Classical Seasonal Decomposition by Moving Averages (Diabetes in CA Example)



Set	NLME MAPE	CSD MAPE
Training	4.75	3.98
Testing	5.03	4.41
COVID-19	10.00	8.00

THANKS

Any Questions?