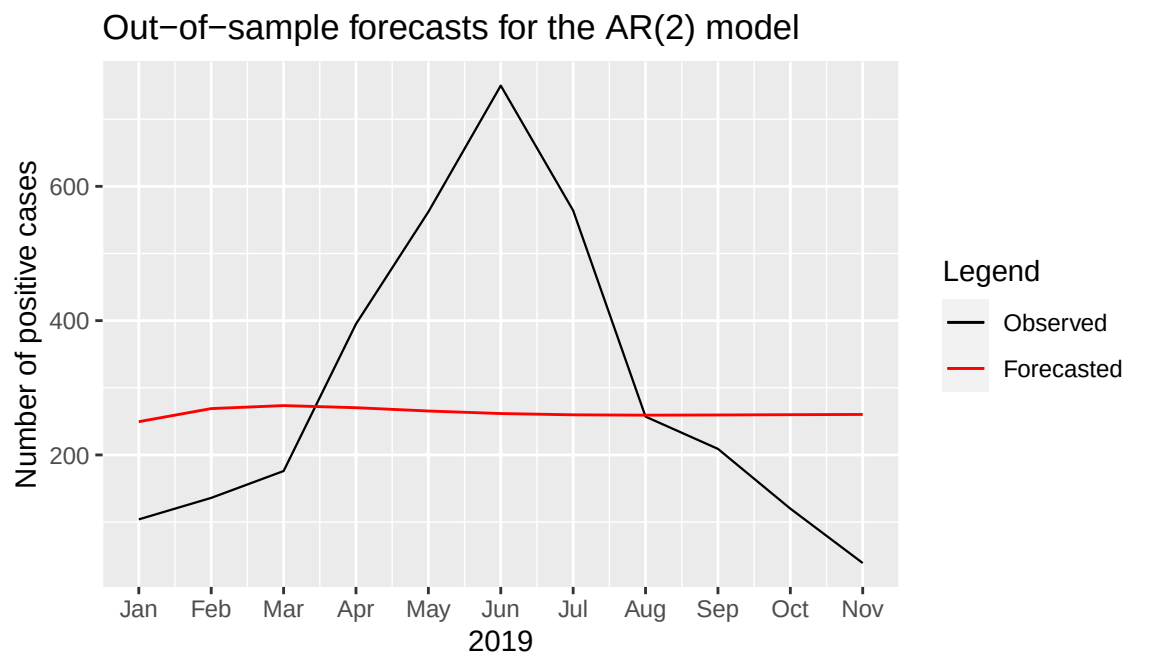
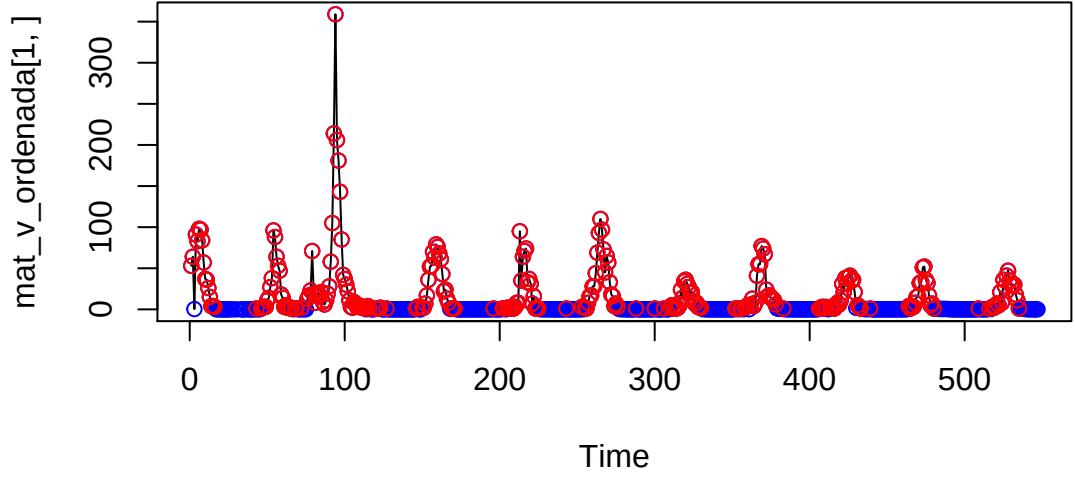
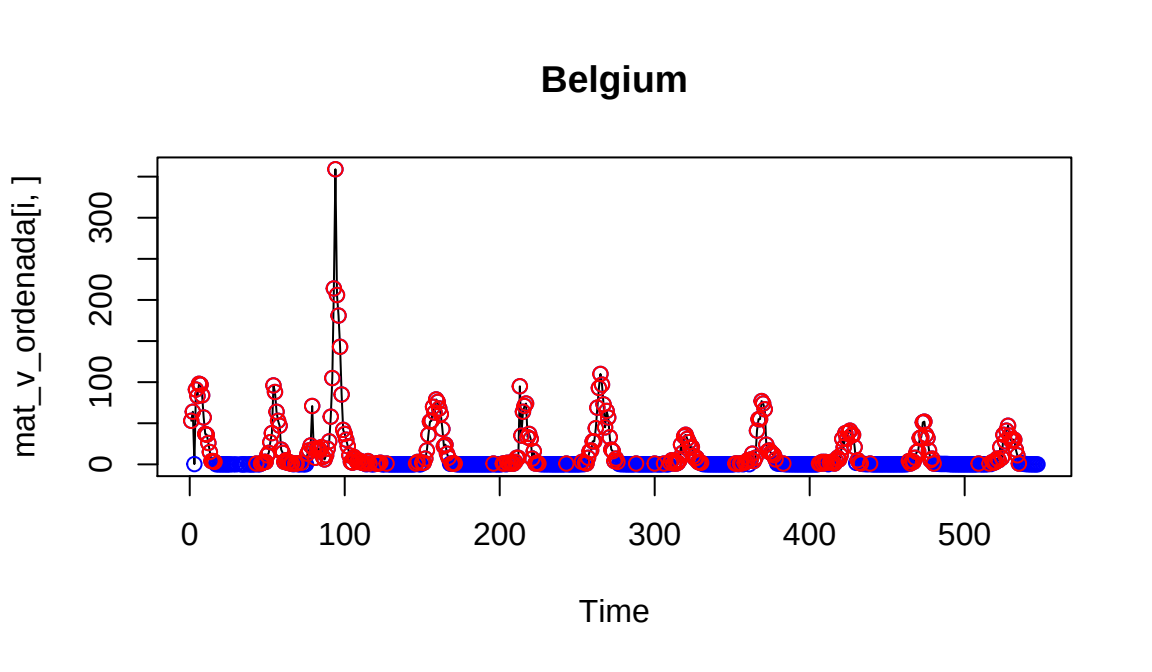
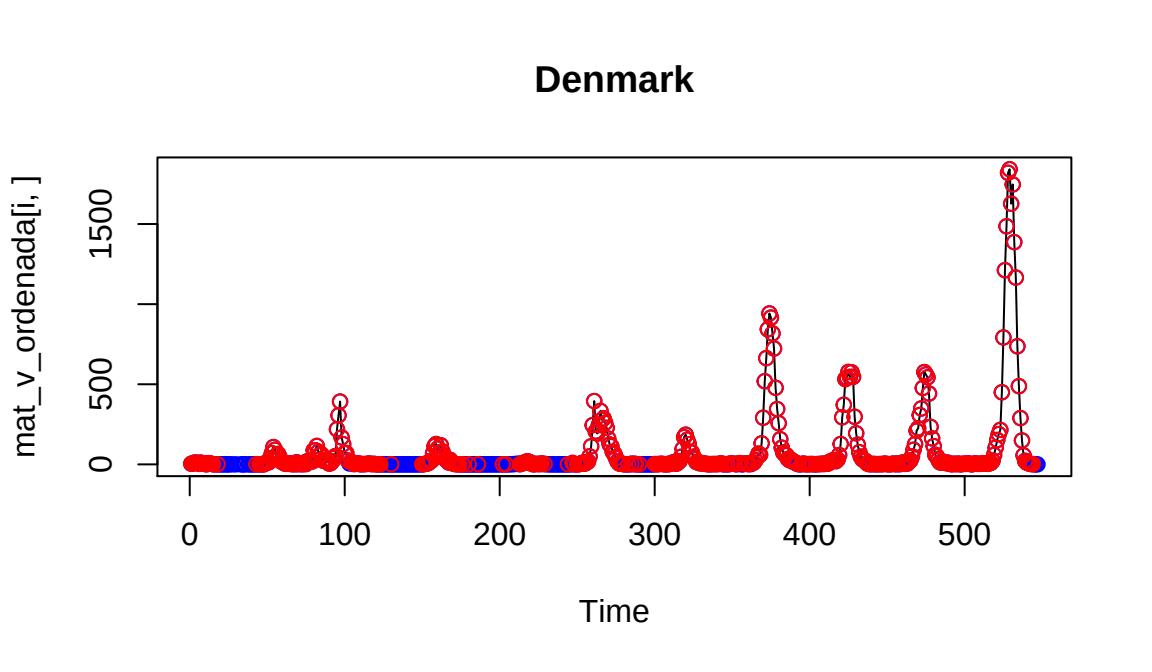
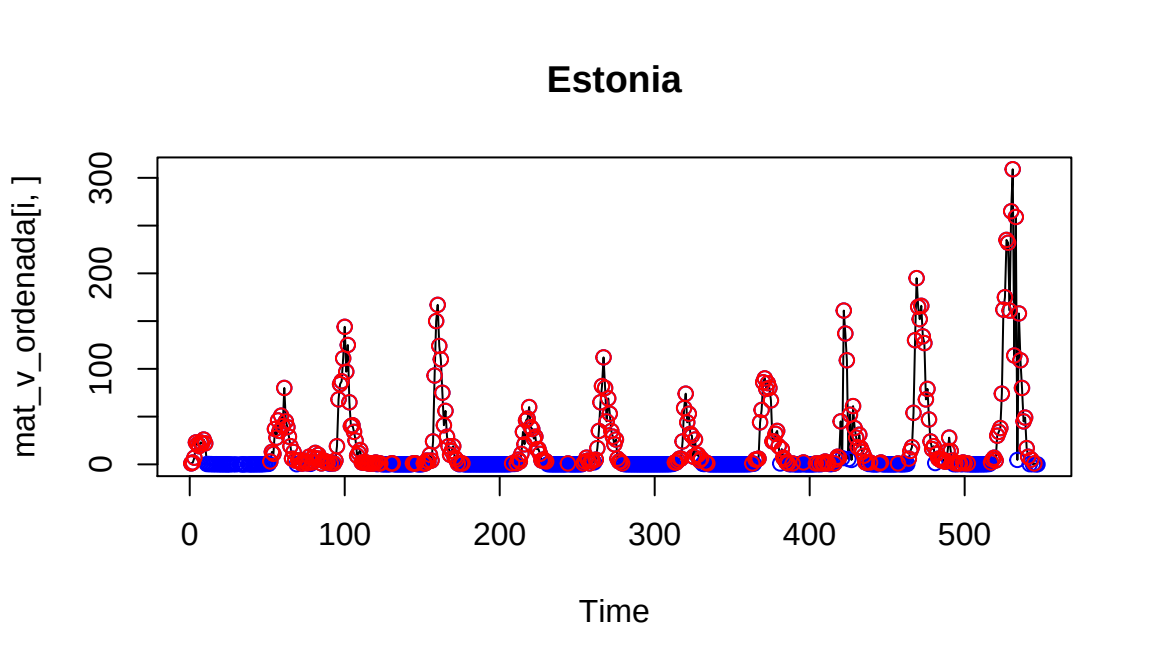
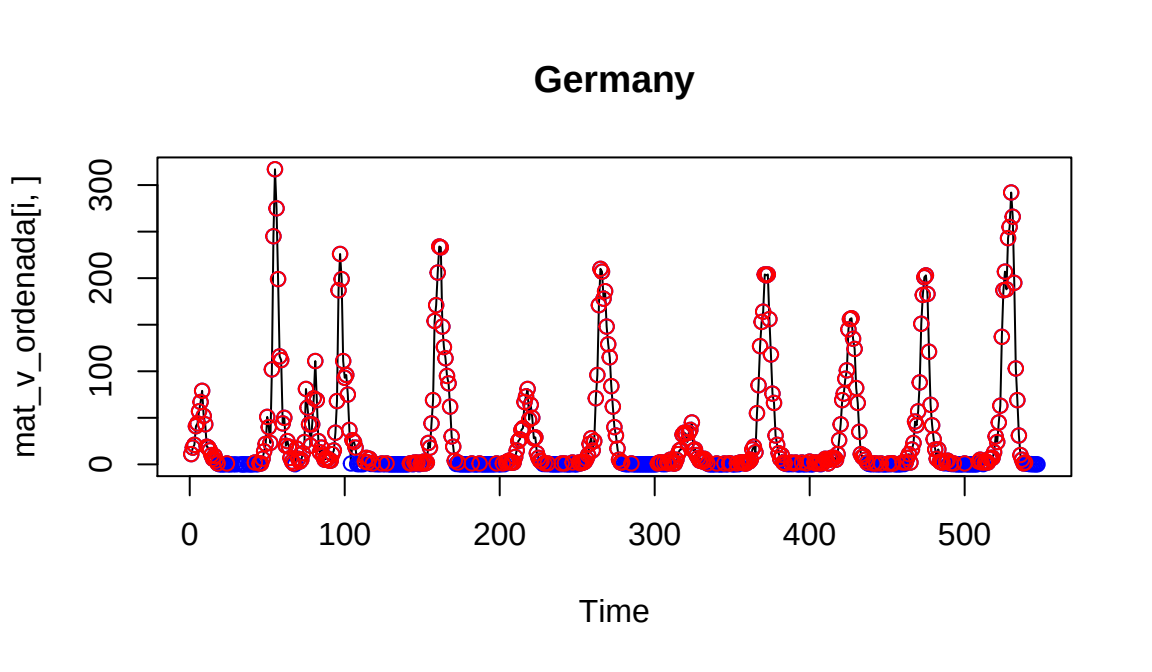

Out−of−sample forecasts for the AR(2) model
Number of positive cases
600
400
200
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
2019
Legend
Observed
Forecasted

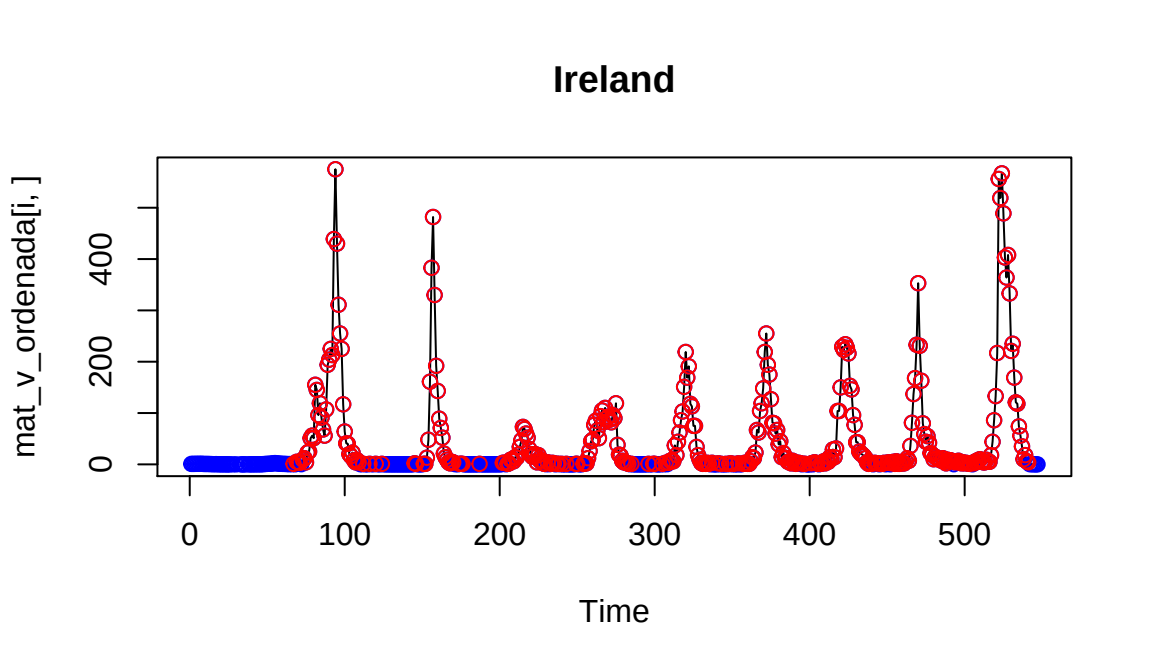

mat_v_ordenada[1,]
0
100
200
300
0
100
200
300
400
500
Time

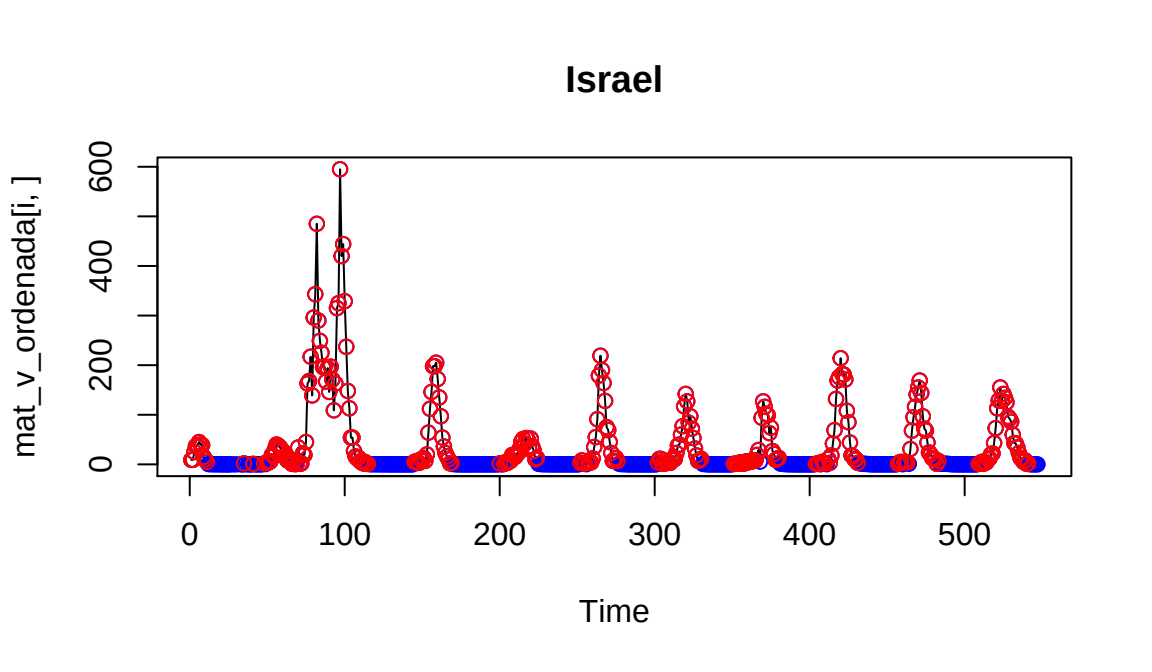

Belgium
mat_v_ordenada[i,]
0
100
200
300
0
100
200
300
400
500
Time

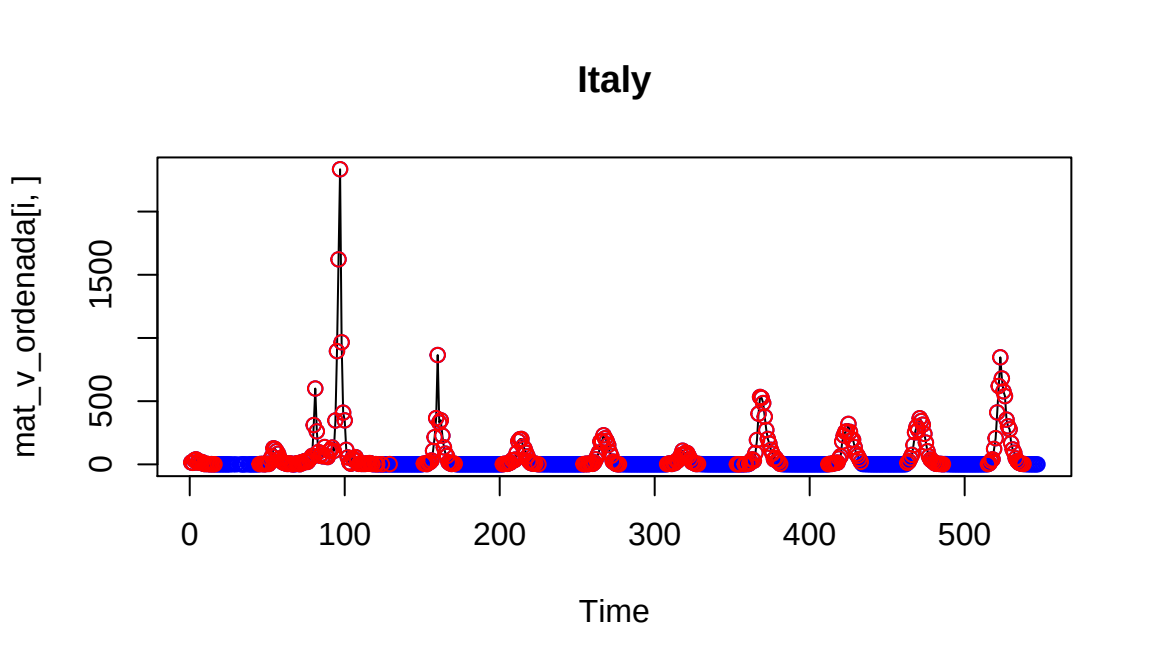

Denmark
mat_v_ordenada[i,]
0
500
1500
0
100
200
300
400
500
Time

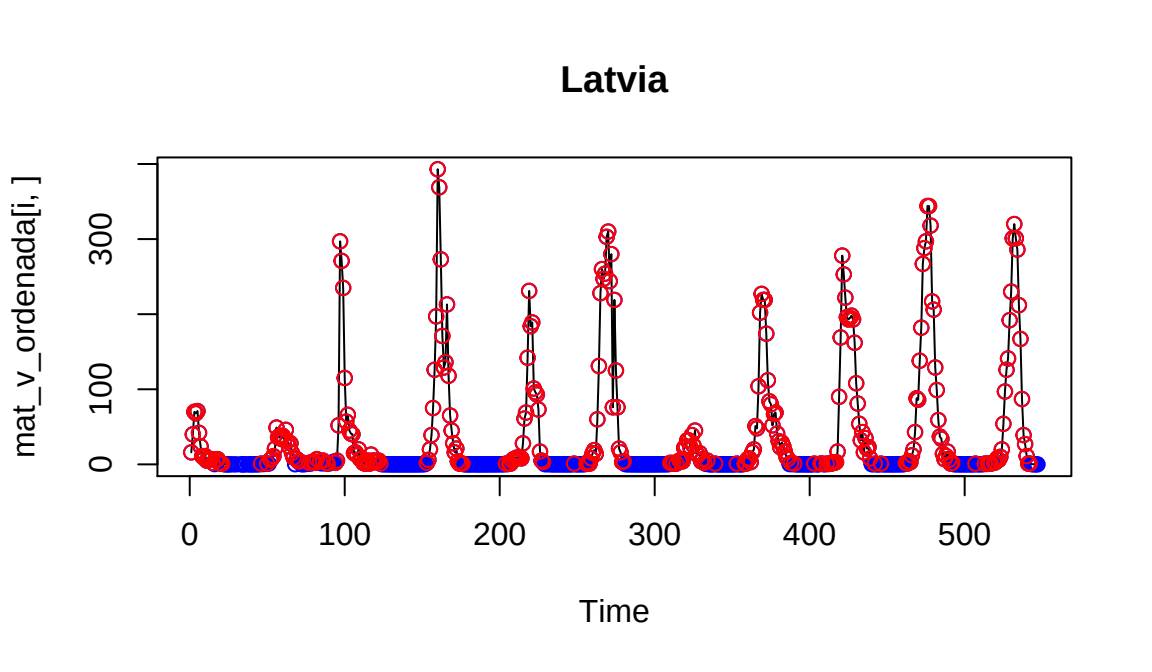

Estonia
mat_v_ordenada[i,]
0
100
200
300
0
100
200
300
400
500
Time

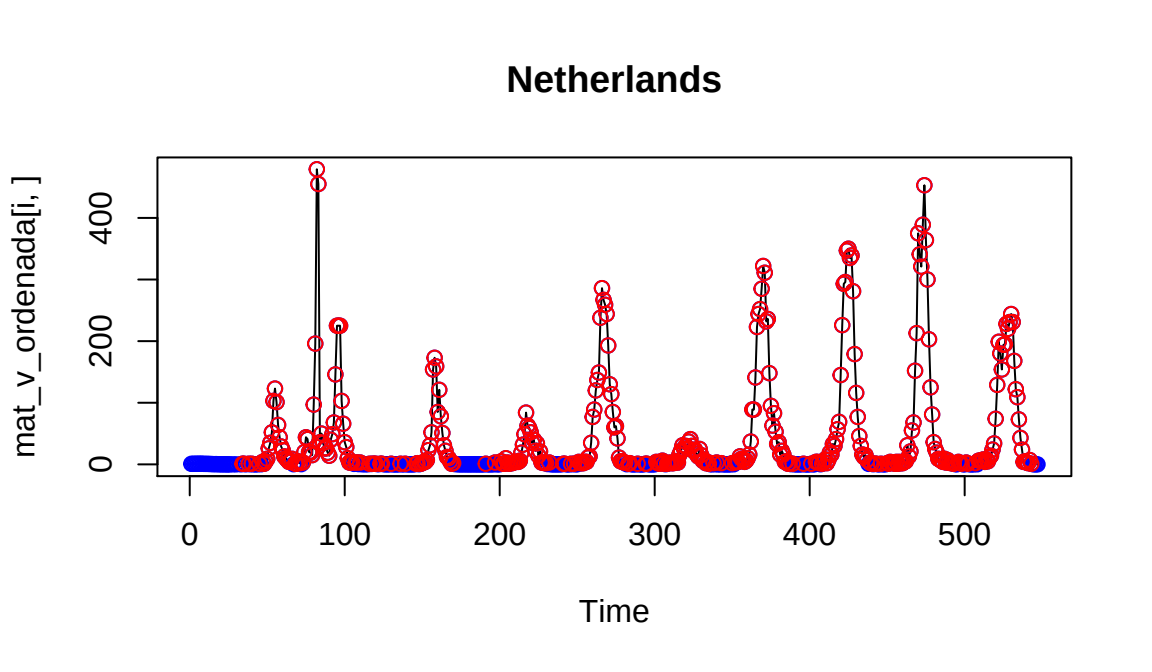

Germany
mat_v_ordenada[i,]
0
100
200
300
0
100
200
300
400
500
Time

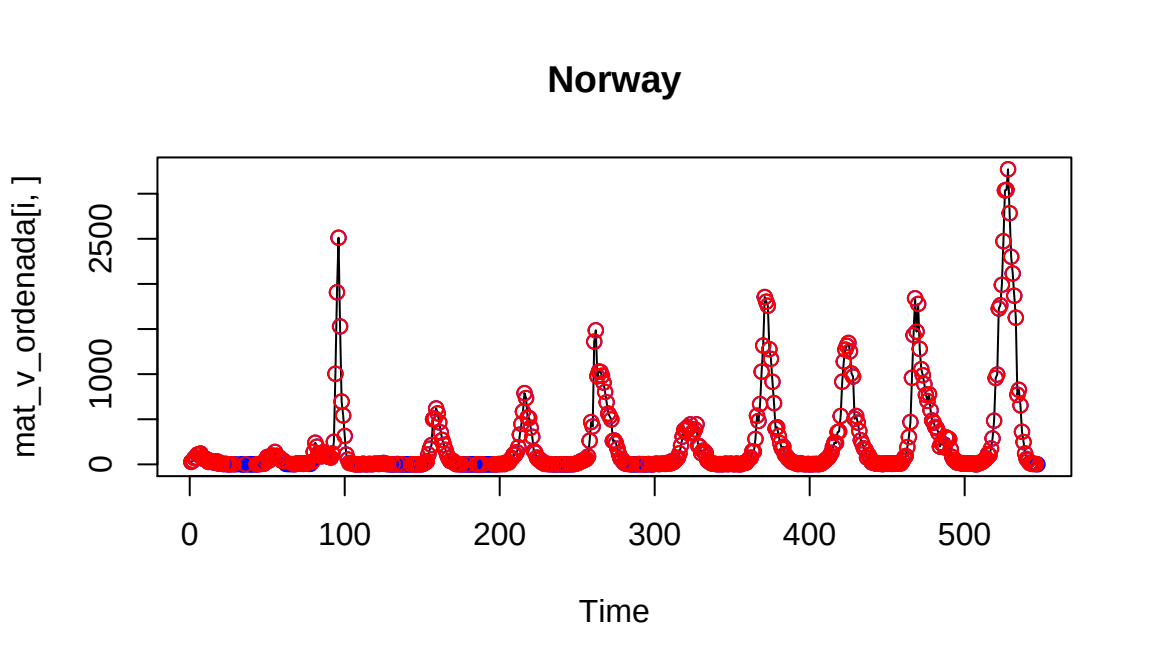

Ireland
mat_v_ordenada[i,]
0
200
400
0
100
200
300
400
500
Time

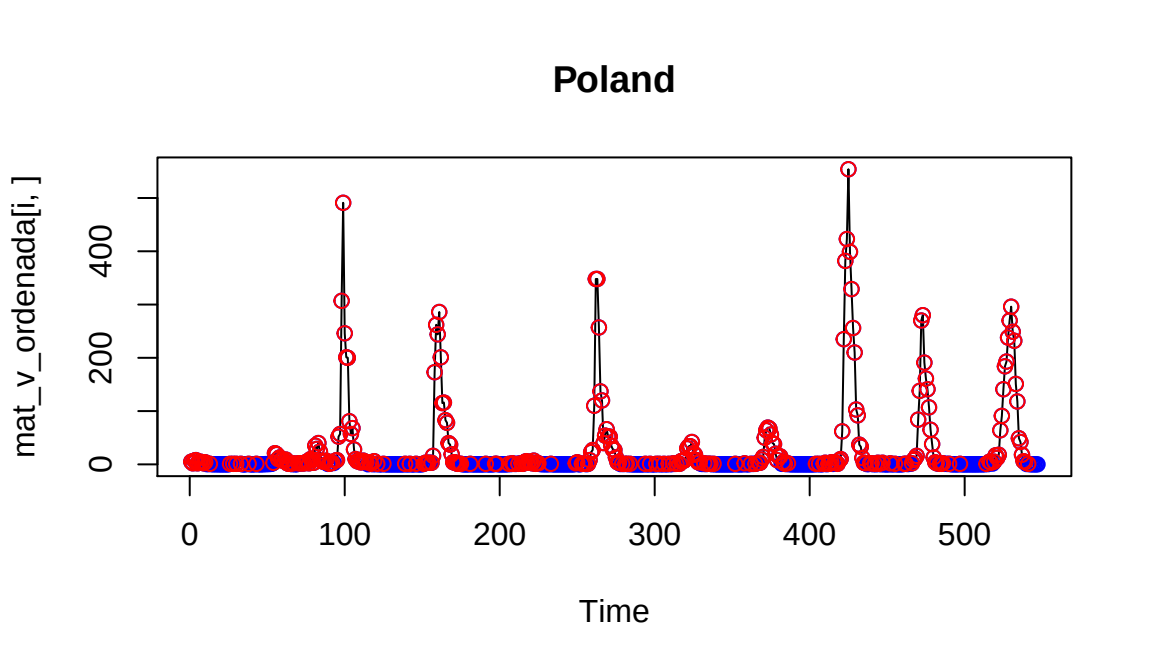

Israel
mat_v_ordenada[i,]
0
200
400
600
0
100
200
300
400
500
Time

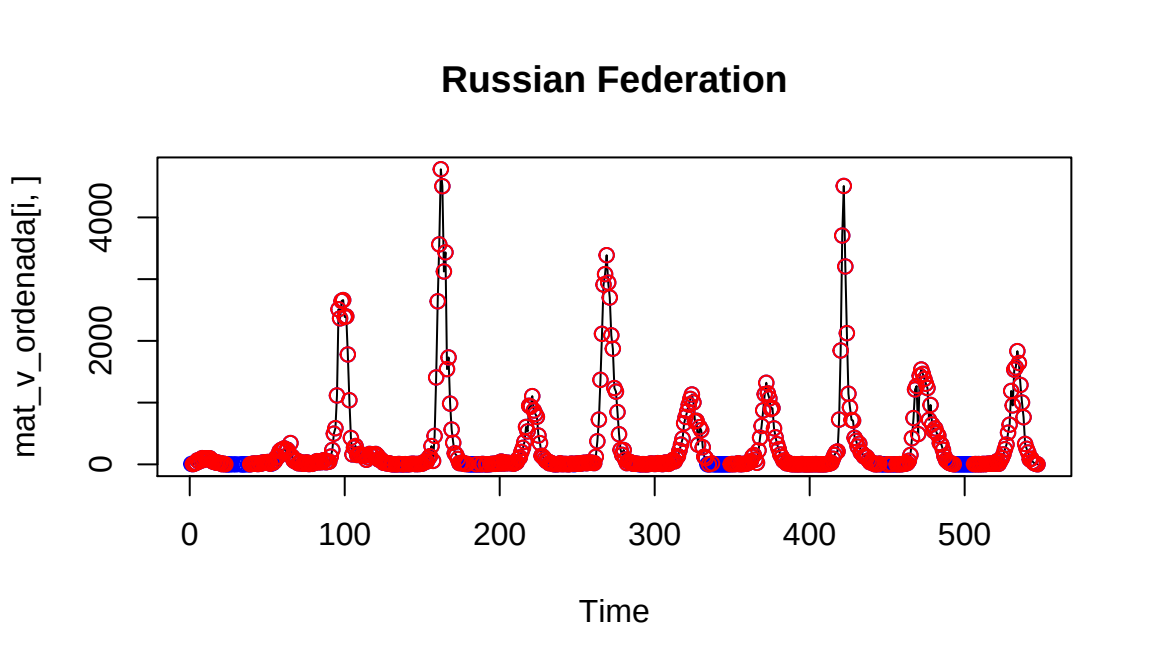

Italy
mat_v_ordenada[i,]
0
500
1500
0
100
200
300
400
500
Time

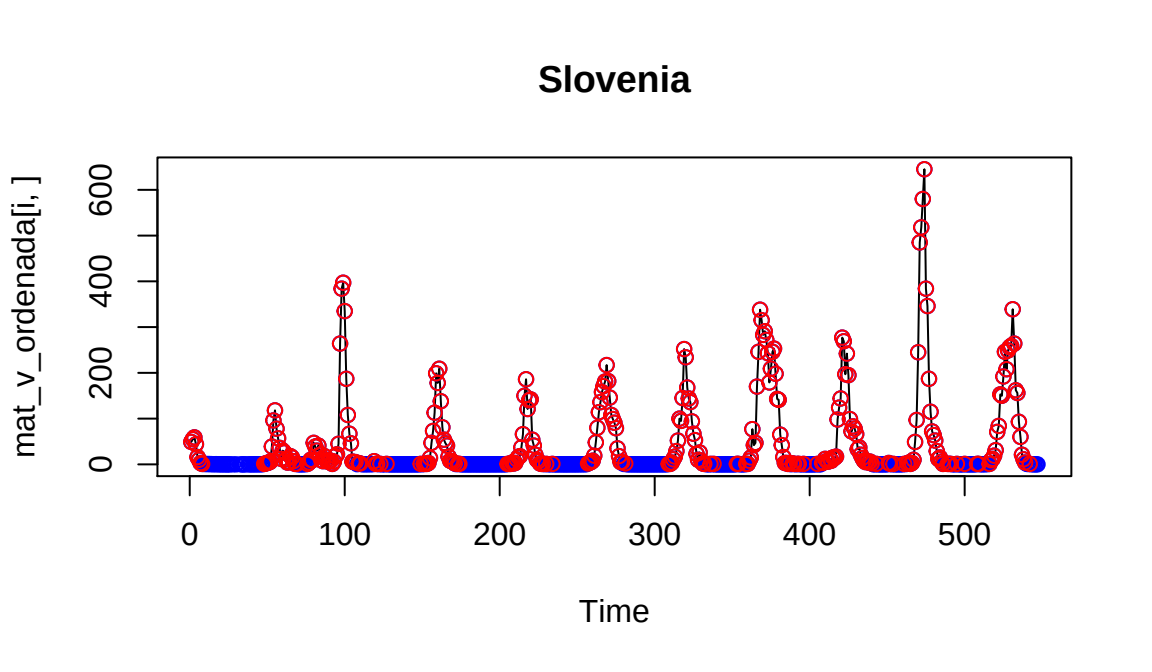

Latvia
mat_v_ordenada[i,]
0
100
300
0
100
200
300
400
500
Time

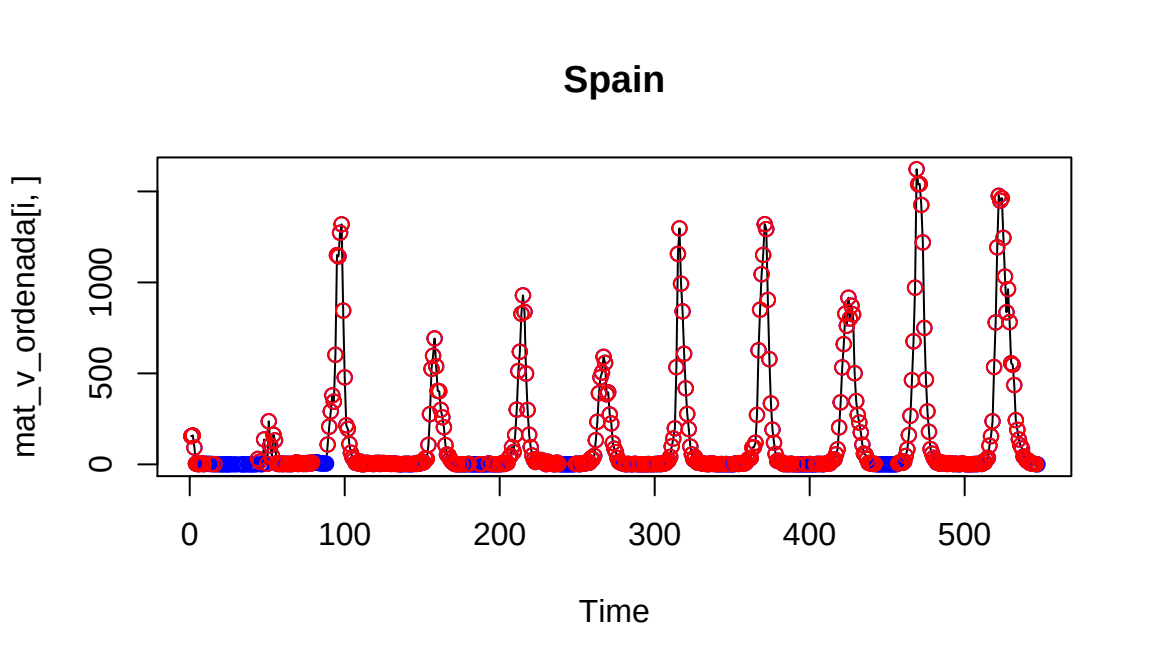

Netherlands
mat_v_ordenada[i,]
0
200
400
0
100
200
300
400
500
Time

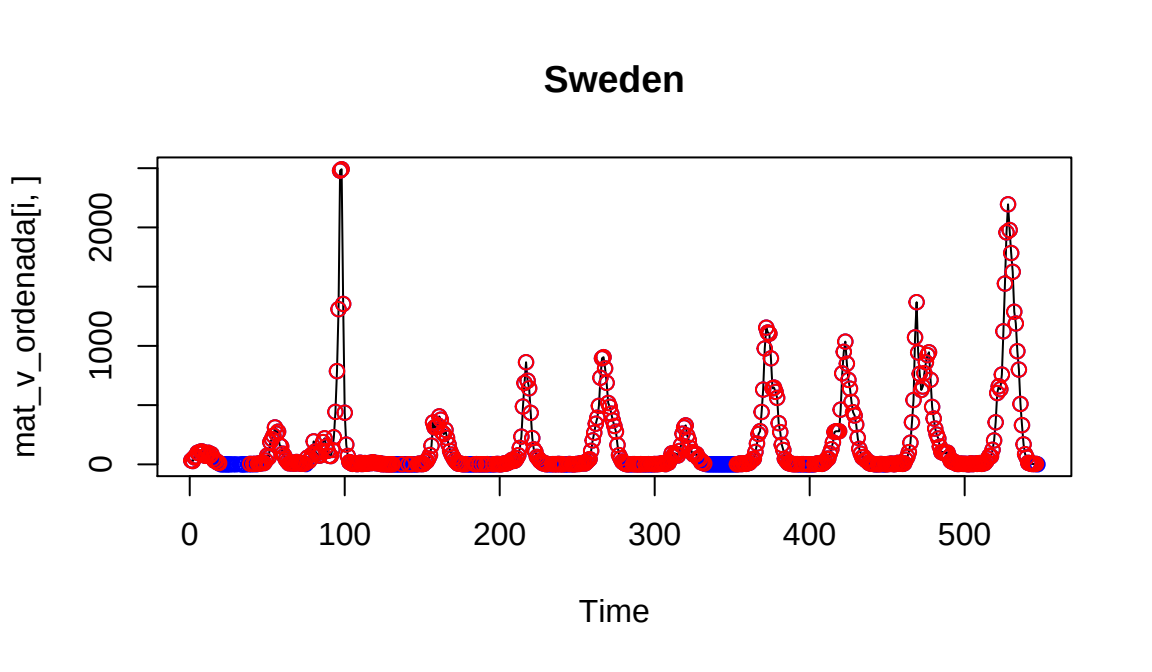

Norway
mat_v_ordenada[i,]
0
1000
2500
0
100
200
300
400
500
Time

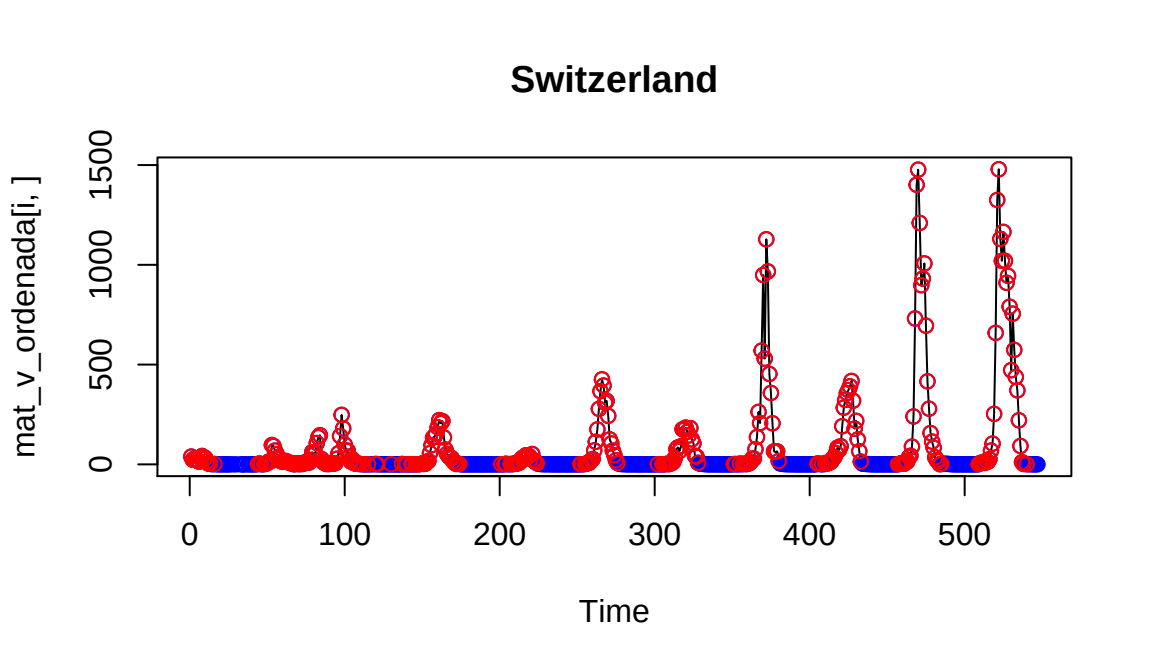

Poland
mat_v_ordenada[i,]
0
200
400
0
100
200
300
400
500
Time

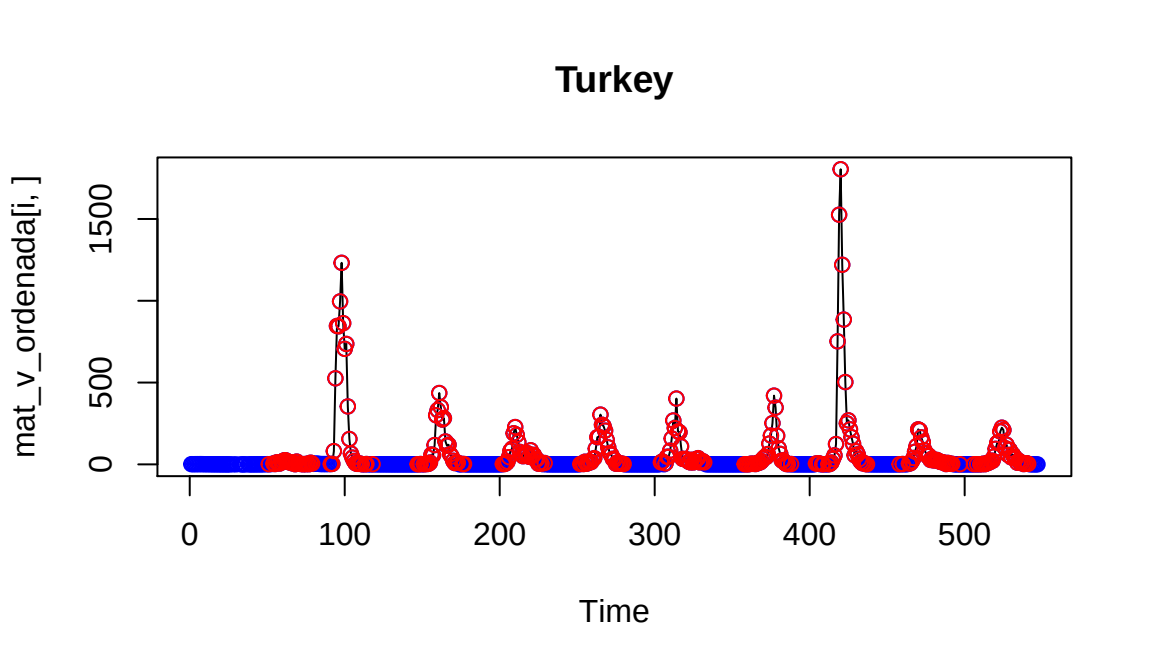

Russian Federation
mat_v_ordenada[i,]
0
2000
4000
0
100
200
300
400
500
Time

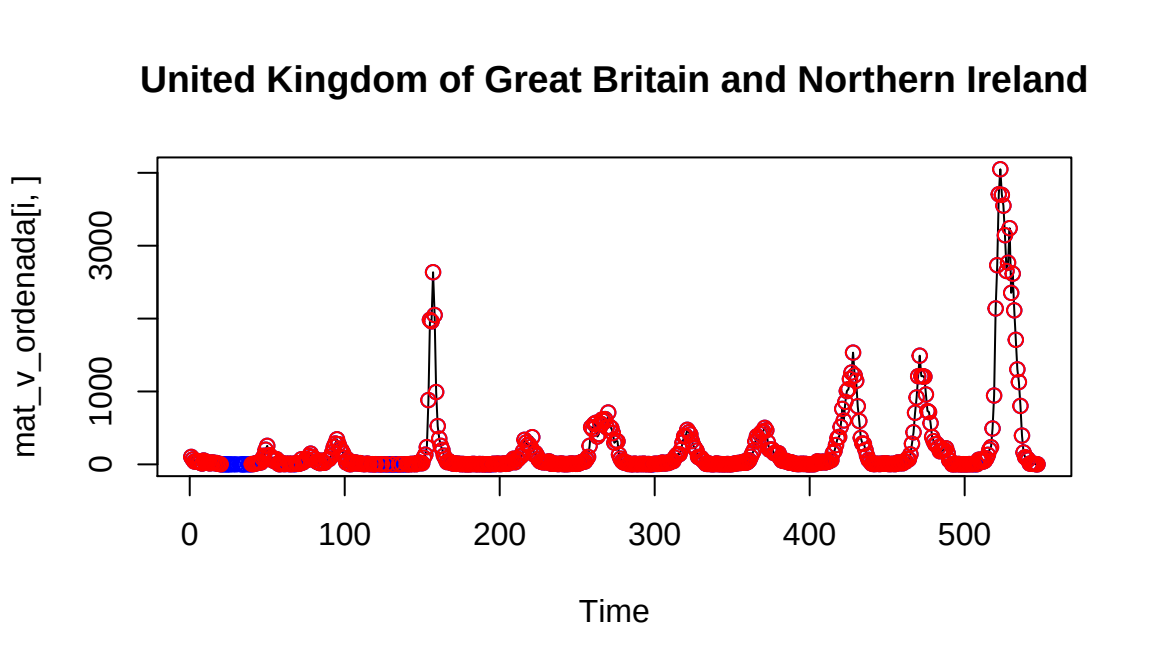

Slovenia
mat_v_ordenada[i,]
0
200
400
600
0
100
200
300
400
500
Time

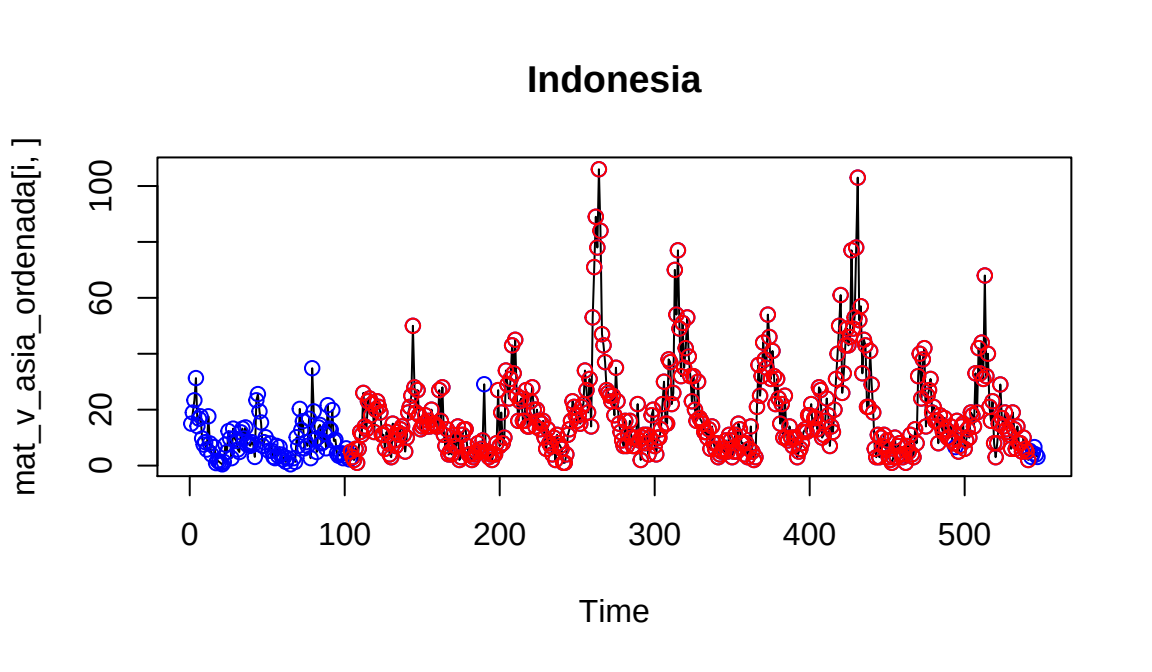

Spain
mat_v_ordenada[i,]
0
500
1000
0
100
200
300
400
500
Time

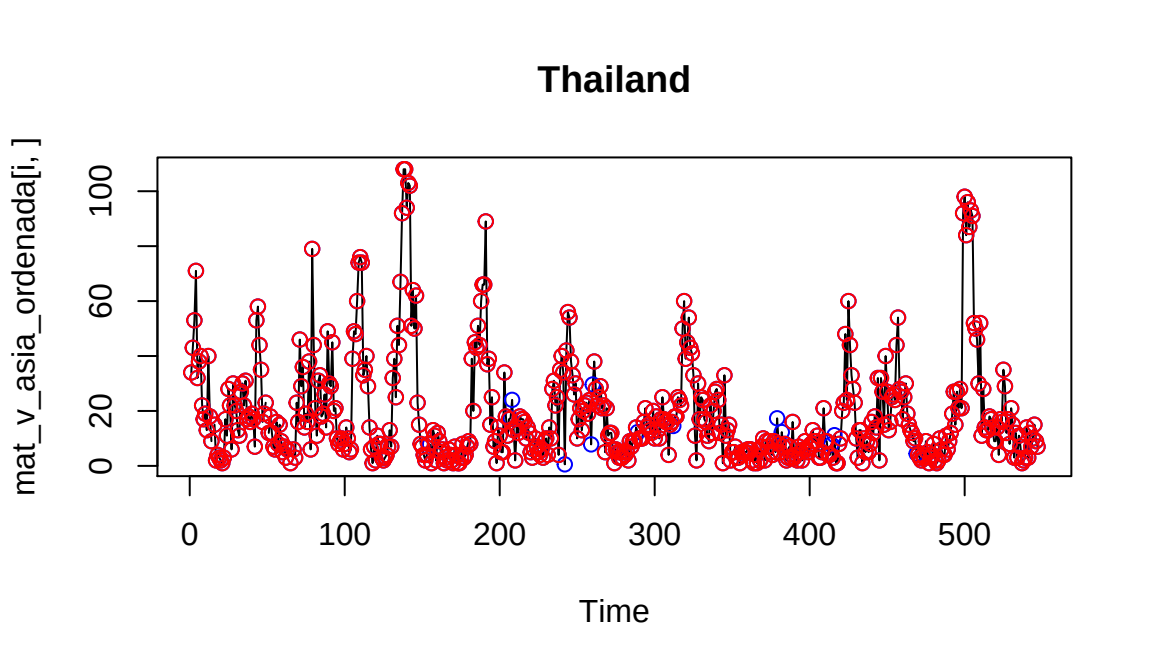

Sweden
mat_v_ordenada[i,]
0
1000
2000
0
100
200
300
400
500
Time

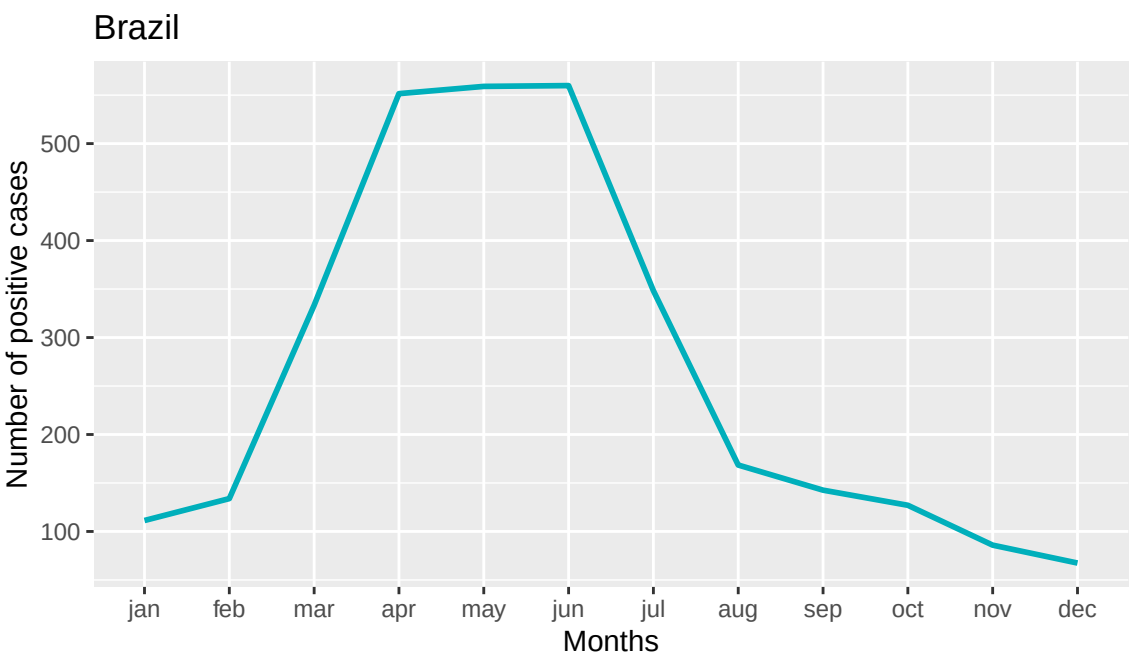

Switzerland
mat_v_ordenada[i,]
0
500
1000
1500
0
100
200
300
400
500
Time

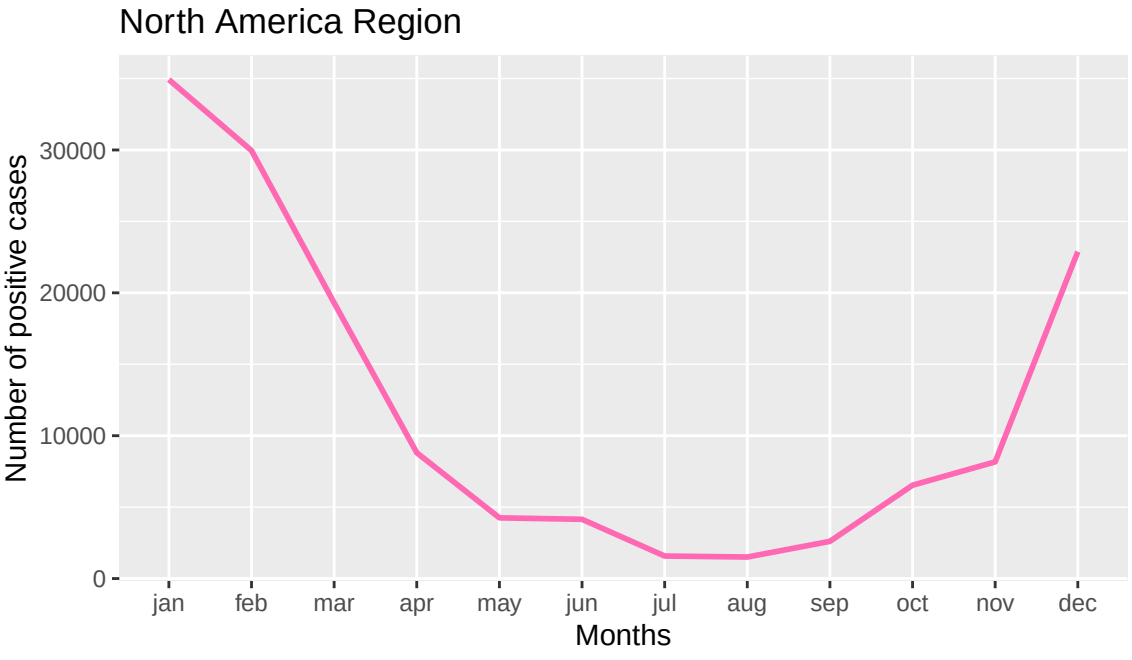

Turkey
mat_v_ordenada[i,]
0
500
1500
0
100
200
300
400
500
Time

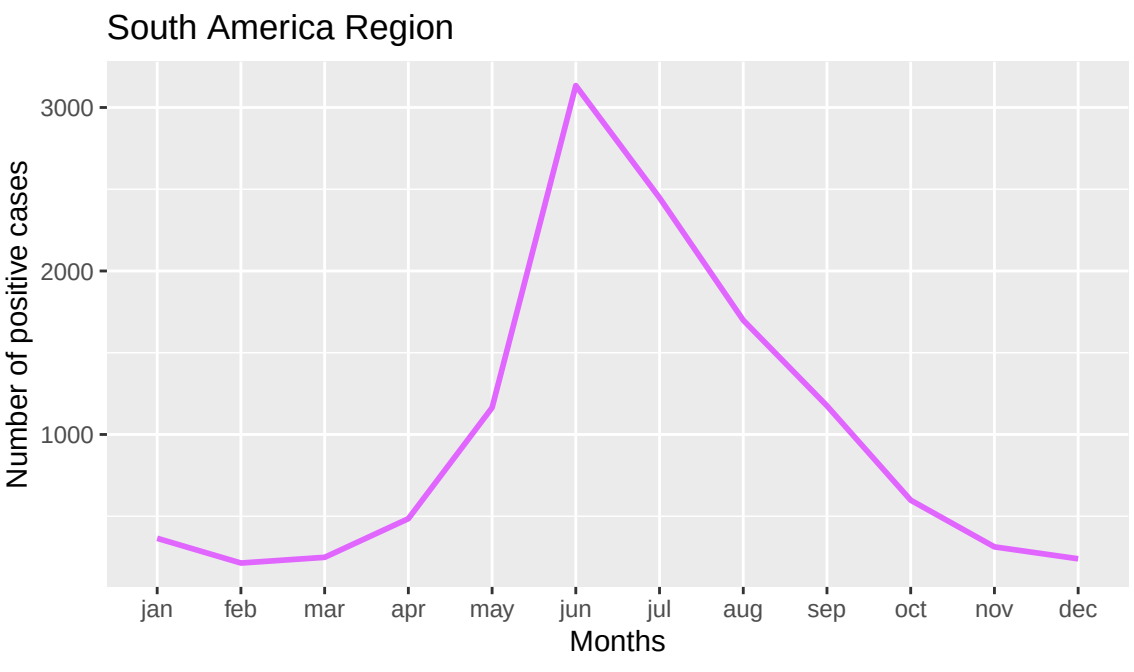

United Kingdom of Great Britain and Northern Ireland
mat_v_ordenada[i,]
0
1000
3000
0
100
200
300
400
500
Time

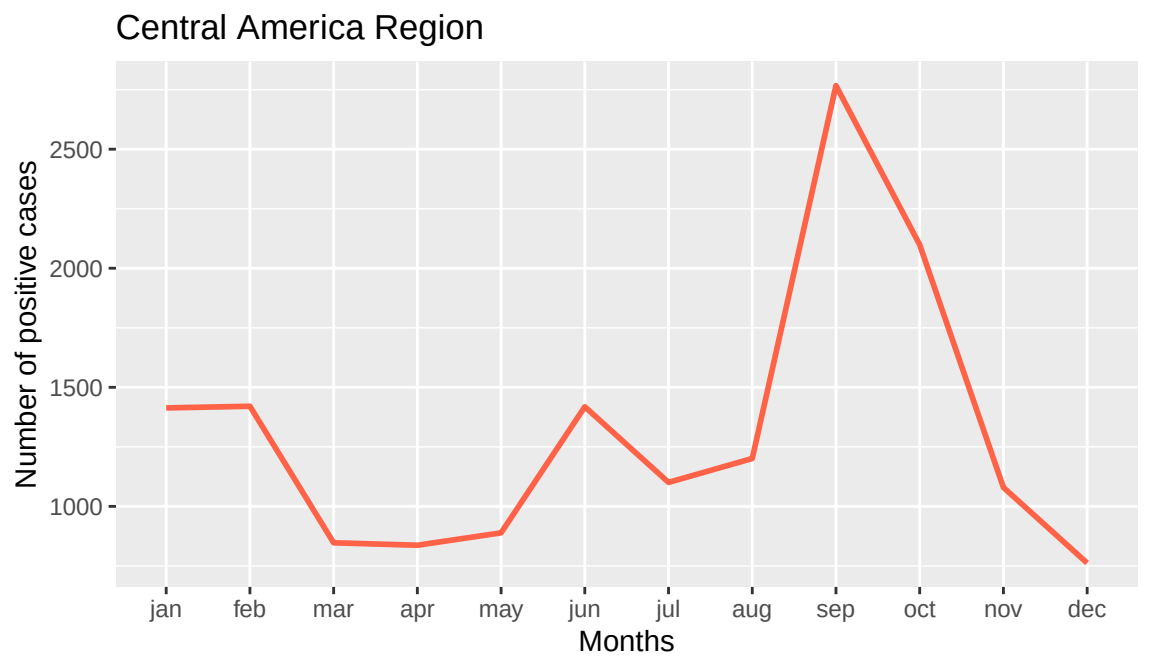

Indonesia
mat_v_asia_ordenada[i,]
0
20
60
100
0
100
200
300
400
500
Time

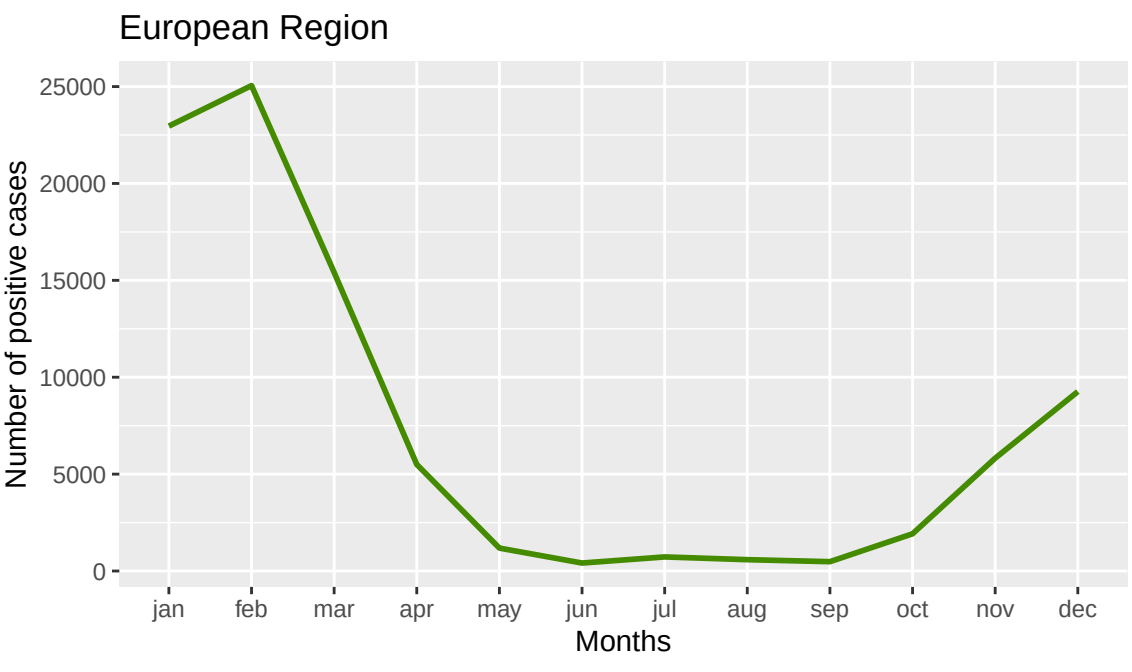

Thailand
mat_v_asia_ordenada[i,]
0
20
60
100
0
100
200
300
400
500
Time

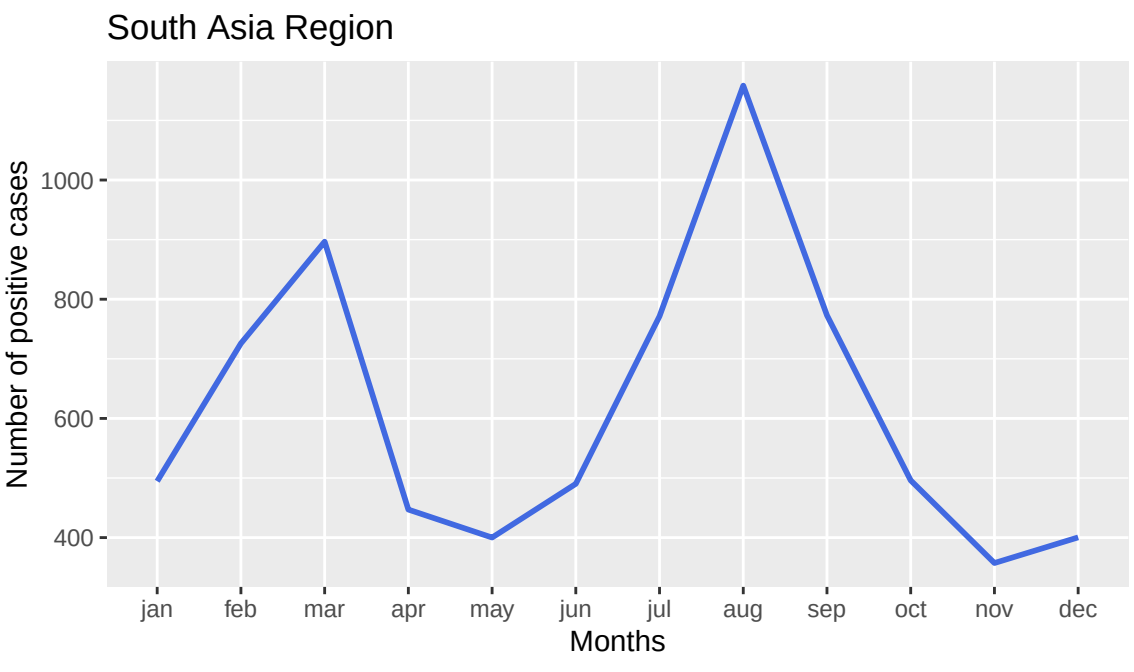

Brazil
Number of positive cases
500
400
300
200
100
jan
feb
mar
apr
may
jun
jul
aug
sep
oct
nov
dec
Months

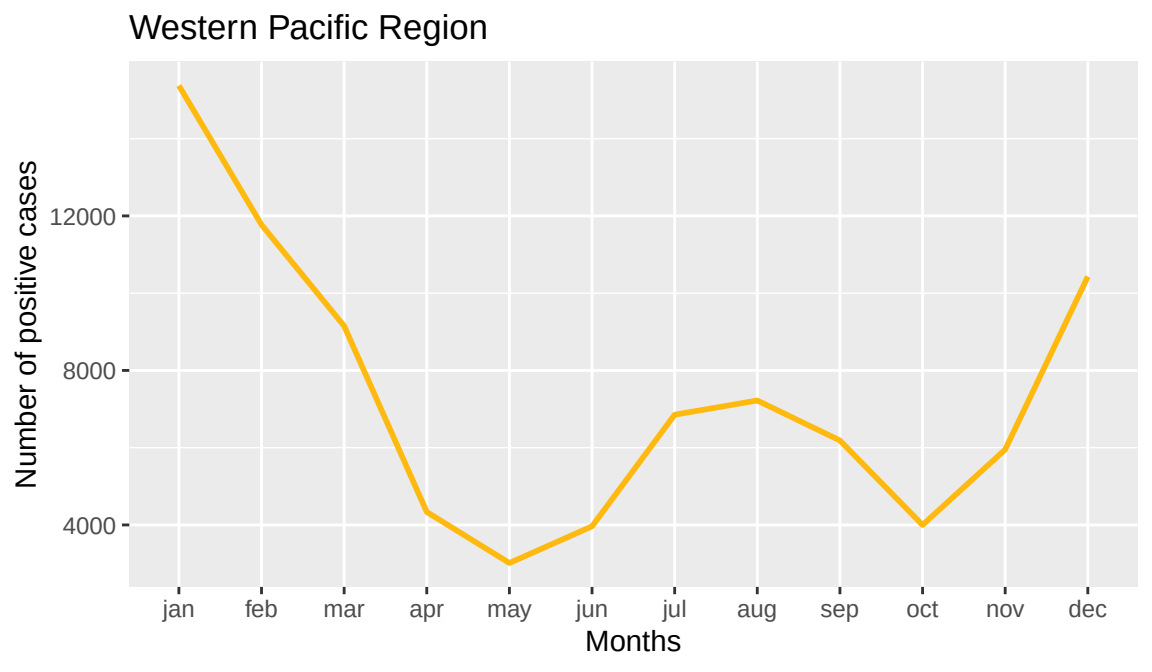

North America Region
Number of positive cases
30000
20000
10000
0
jan
feb
mar
apr
may
jun
jul
aug
sep
oct
nov
dec
Months

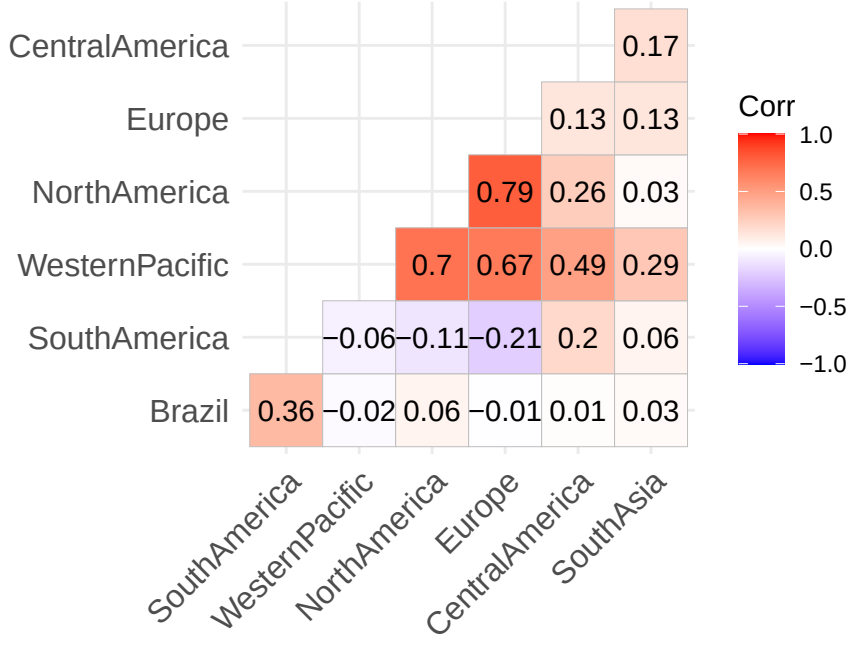

South America Region
Number of positive cases
3000
2000
1000
jan
feb
mar
apr
may
jun
jul
aug
sep
oct
nov
dec
Months

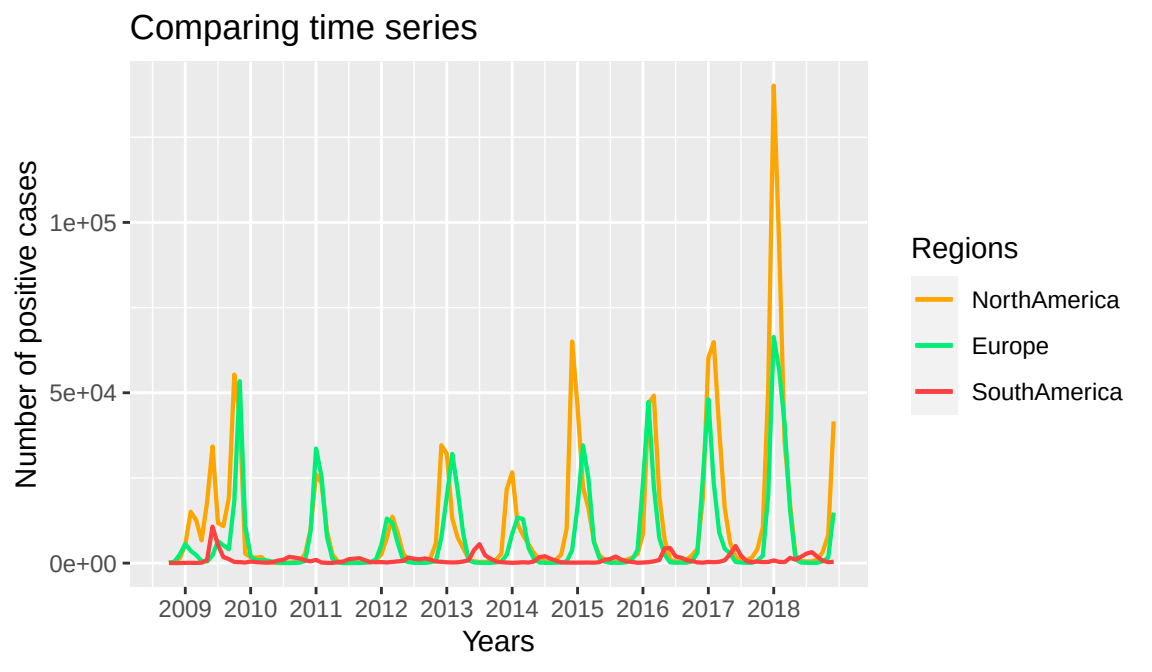

Central America Region
Number of positive cases
2500
2000
1500
1000
jan
feb
mar
apr
may
jun
jul
aug
sep
oct
nov
dec
Months

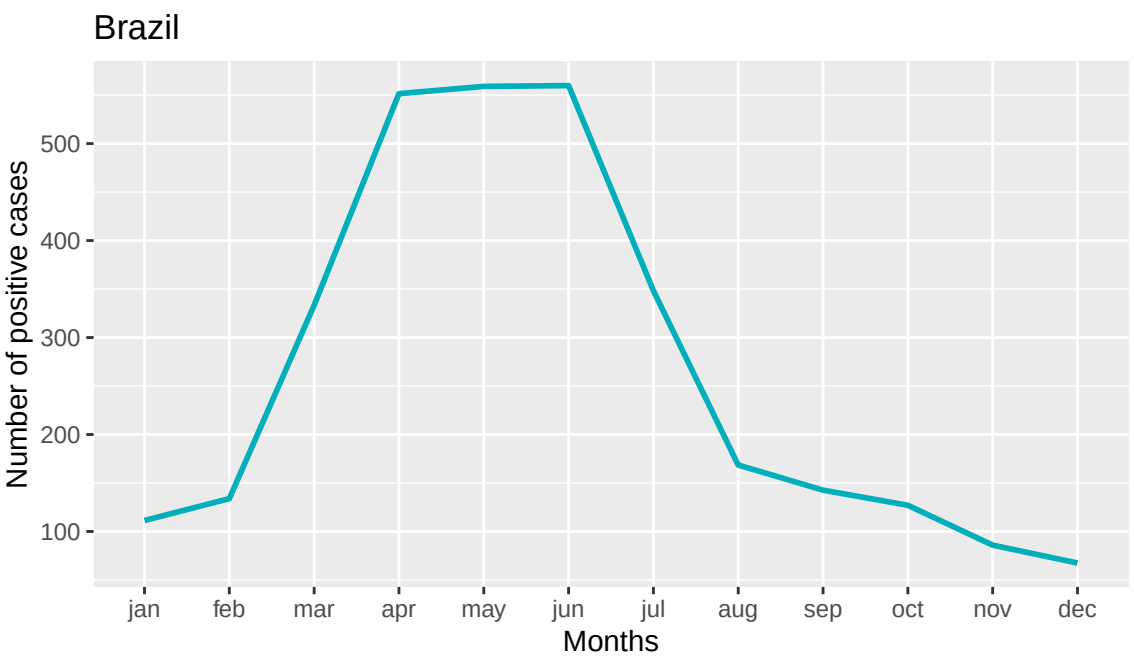

European Region
Number of positive cases
25000
20000
15000
10000
5000
0
jan
feb
mar
apr
may
jun
jul
aug
sep
oct
nov
dec
Months

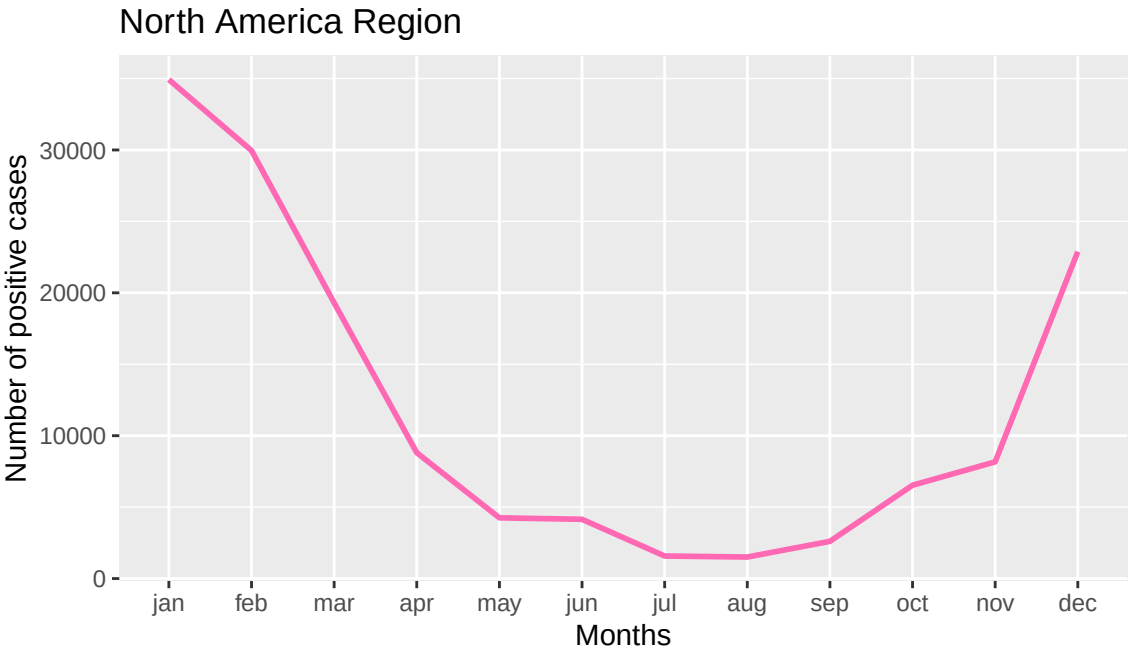

South Asia Region
Number of positive cases
1000
800
600
400
jan
feb
mar
apr
may
jun
jul
aug
sep
oct
nov
dec
Months

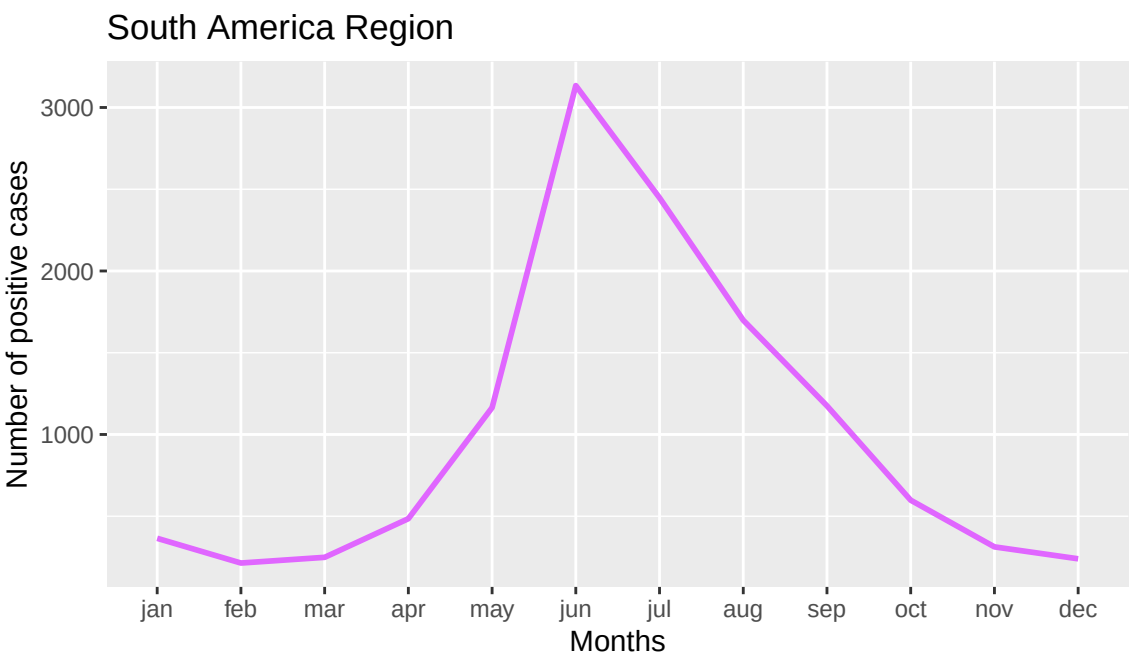

Western Pacific Region
Number of positive cases
12000
8000
4000
jan
feb
mar
apr
may
jun
jul
aug
sep
oct
nov
dec
Months

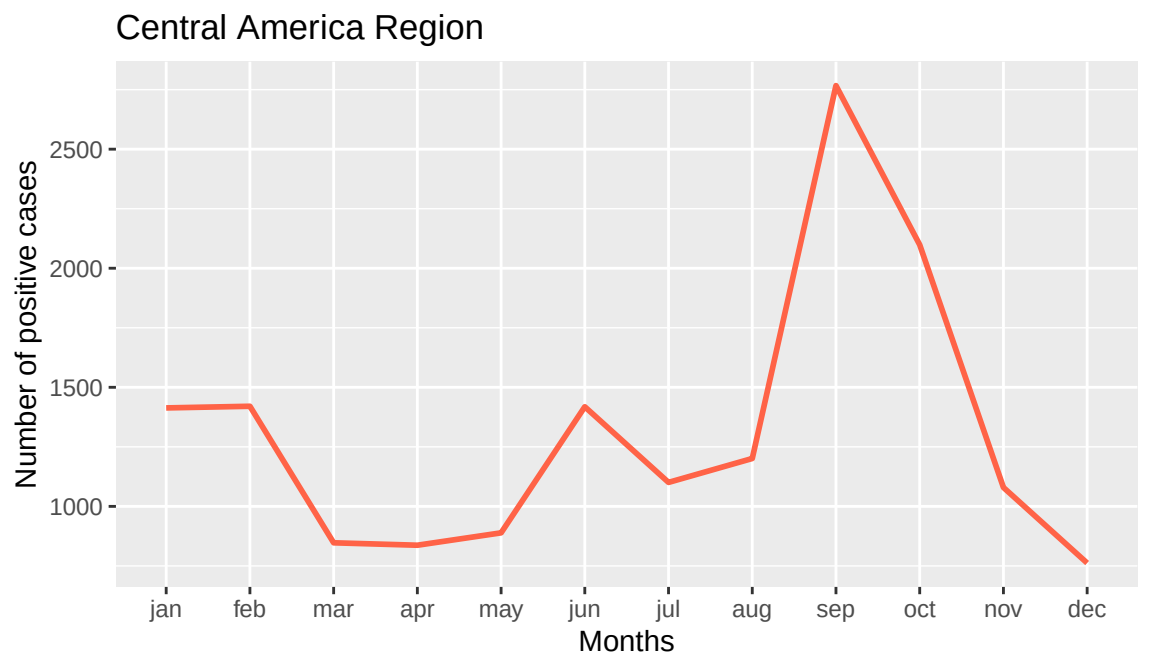

CentralAmerica
Europe
NorthAmerica
WesternPacific
SouthAmerica
Brazil
0.17
0.13 0.13
0.79 0.26 0.03
0.7 0.67 0.49 0.29
−0.06 −0.11 −0.21 0.2 0.06
0.36 −0.02 0.06 −0.01 0.01 0.03
SouthAmerica
WesternPacific
NorthAmerica
Europe
CentralAmerica
SouthAsia
Corr
1.0
0.5
0.0
−0.5
−1.0

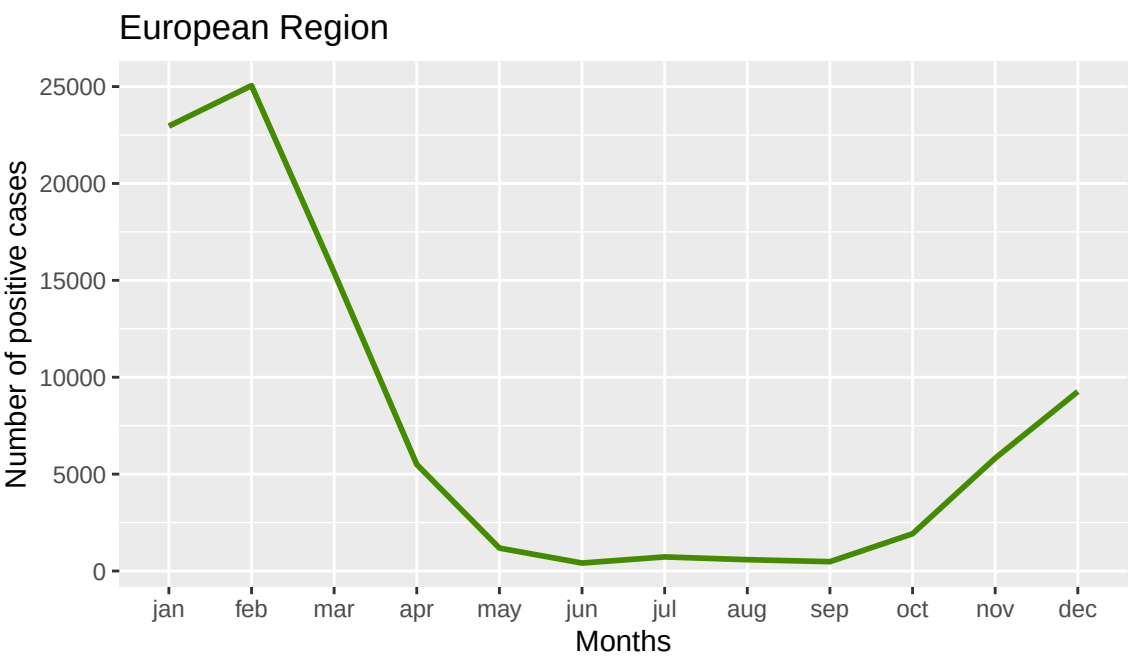

Comparing time series
Number of positive cases
1e+05
5e+04
0e+00
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
Years
Regions
NorthAmerica
Europe
SouthAmerica

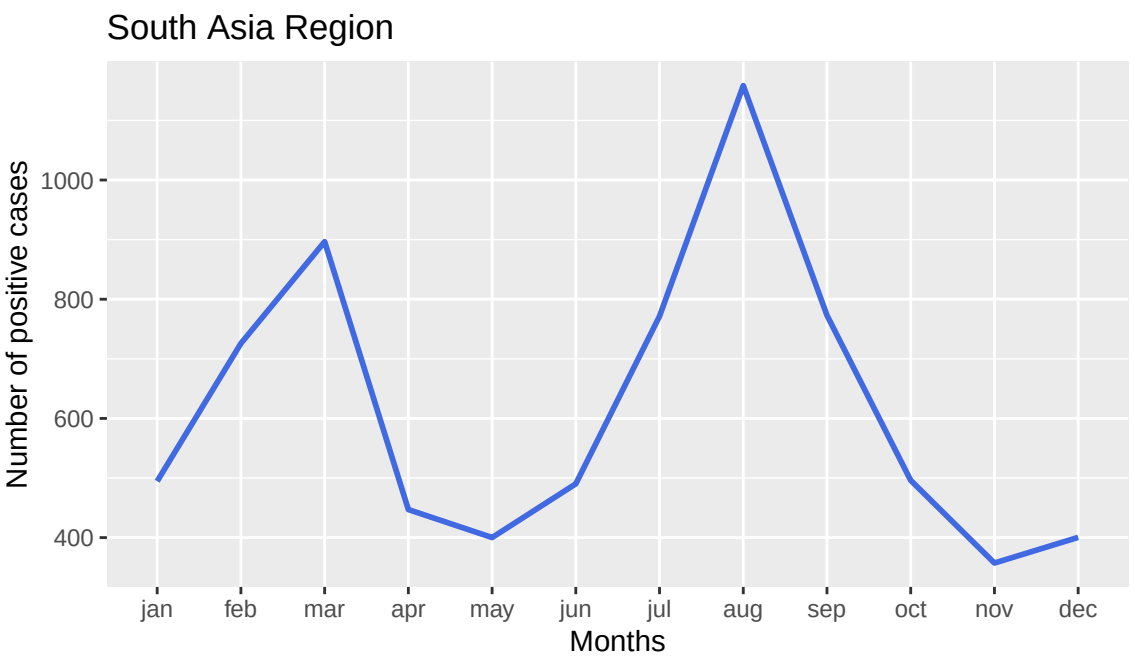

Brazil
Number of positive cases
500
400
300
200
100
jan
feb
mar
apr
may
jun
jul
aug
sep
oct
nov
dec
Months

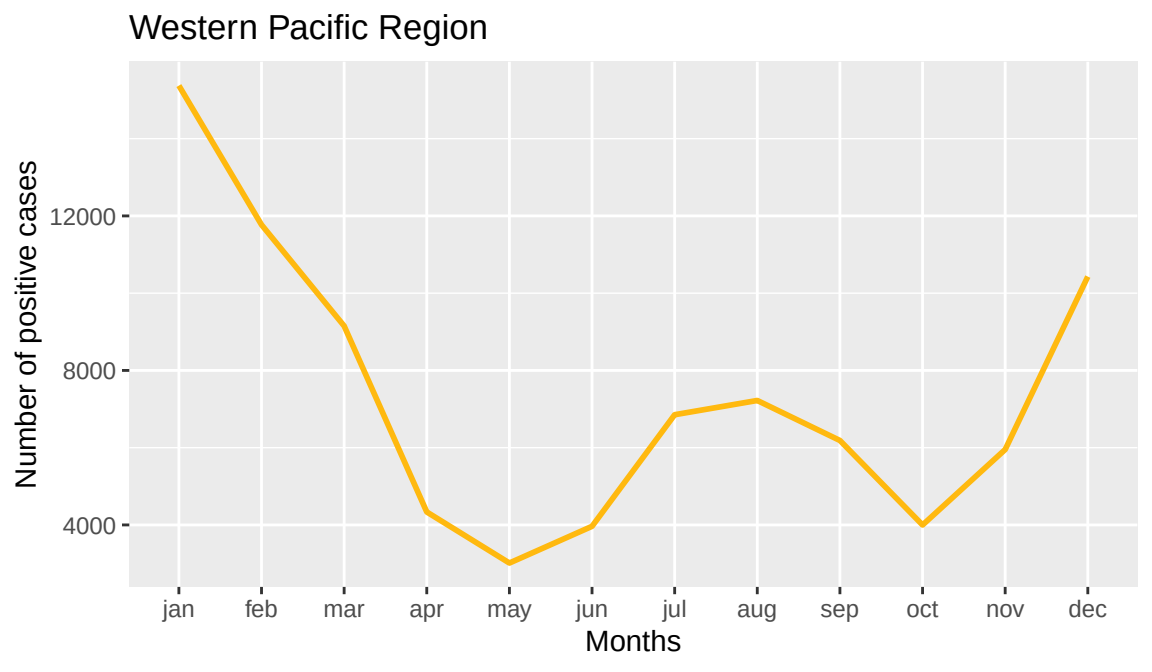

North America Region
Number of positive cases
30000
20000
10000
0
jan
feb
mar
apr
may
jun
jul
aug
sep
oct
nov
dec
Months

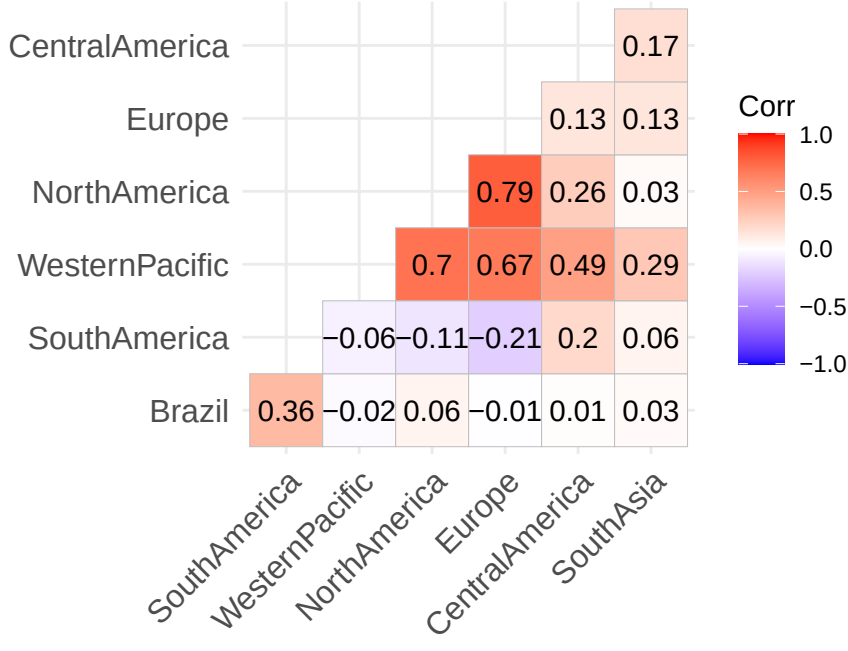

South America Region
Number of positive cases
3000
2000
1000
jan
feb
mar
apr
may
jun
jul
aug
sep
oct
nov
dec
Months

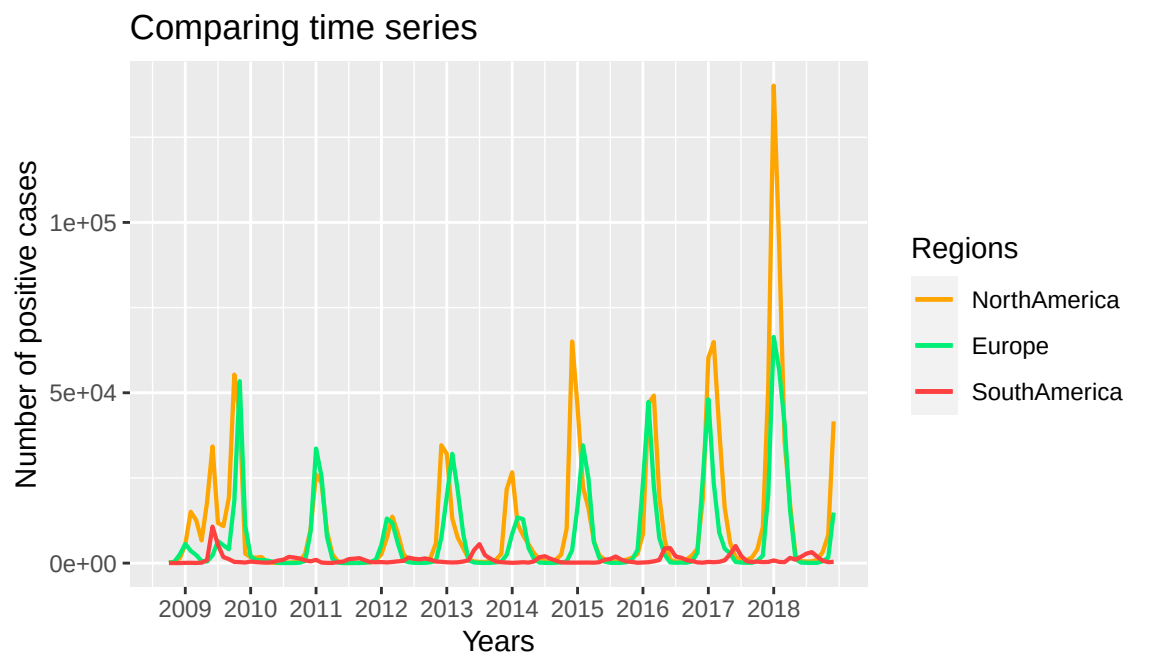

Central America Region
Number of positive cases
2500
2000
1500
1000
jan
feb
mar
apr
may
jun
jul
aug
sep
oct
nov
dec
Months

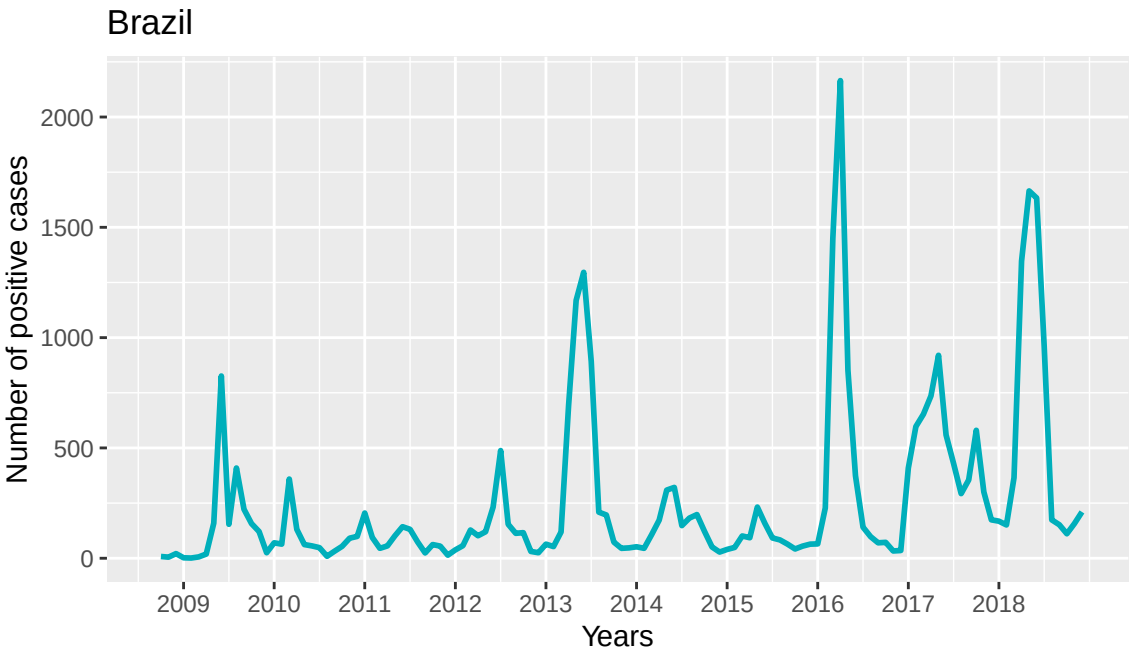

European Region
Number of positive cases
25000
20000
15000
10000
5000
0
jan
feb
mar
apr
may
jun
jul
aug
sep
oct
nov
dec
Months

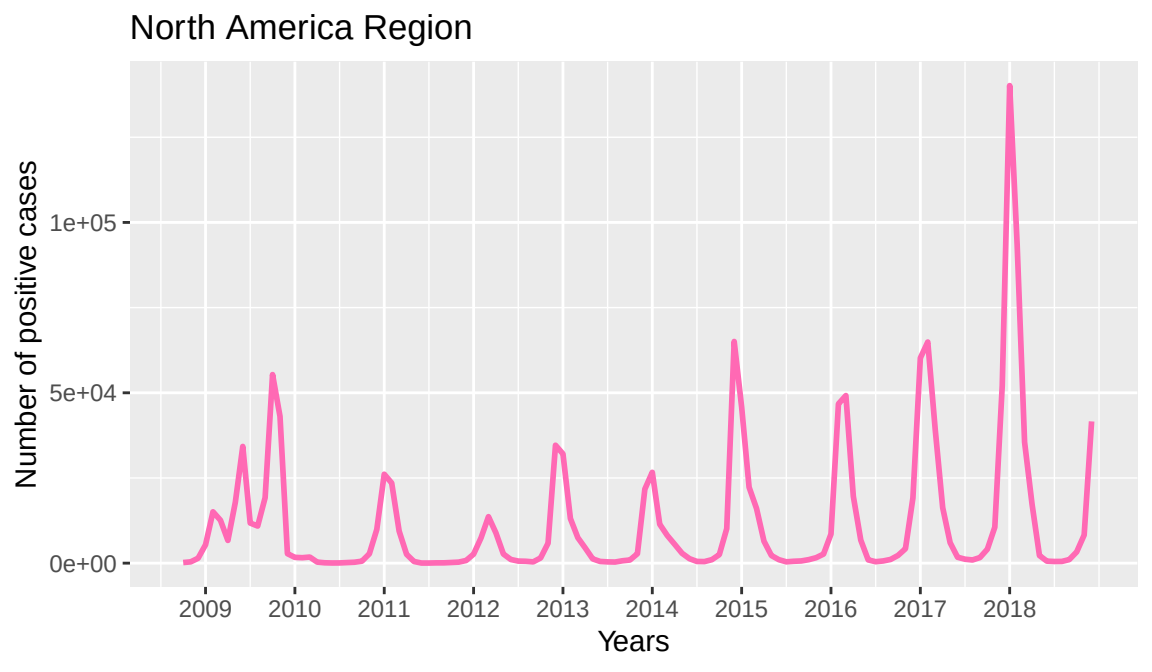

South Asia Region
Number of positive cases
1000
800
600
400
jan
feb
mar
apr
may
jun
jul
aug
sep
oct
nov
dec
Months

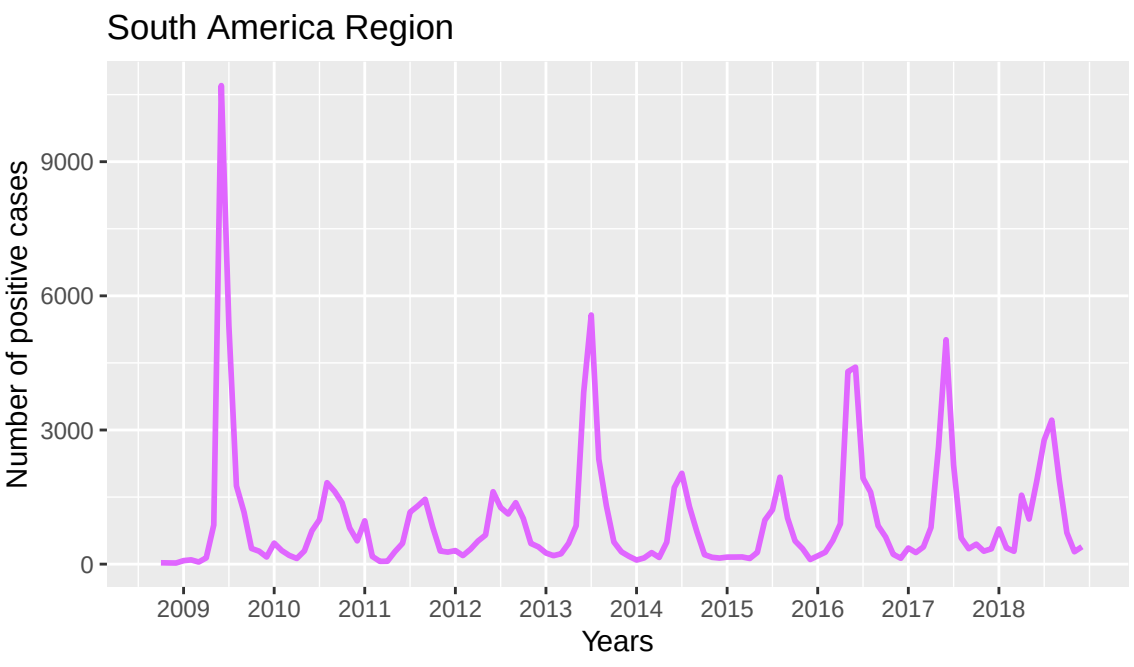

Western Pacific Region
Number of positive cases
12000
8000
4000
jan
feb
mar
apr
may
jun
jul
aug
sep
oct
nov
dec
Months

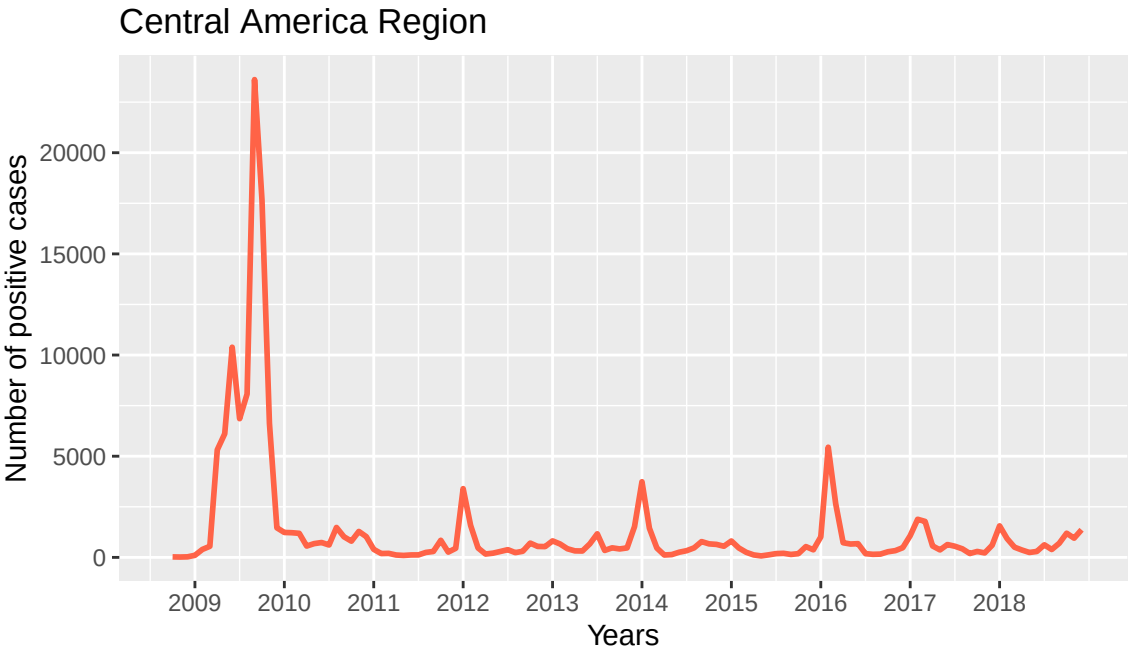

CentralAmerica
Europe
NorthAmerica
WesternPacific
SouthAmerica
Brazil
0.17
0.13 0.13
0.79 0.26 0.03
0.7 0.67 0.49 0.29
−0.06 −0.11 −0.21 0.2 0.06
0.36 −0.02 0.06 −0.01 0.01 0.03
SouthAmerica
WesternPacific
NorthAmerica
Europe
CentralAmerica
SouthAsia
Corr
1.0
0.5
0.0
−0.5
−1.0

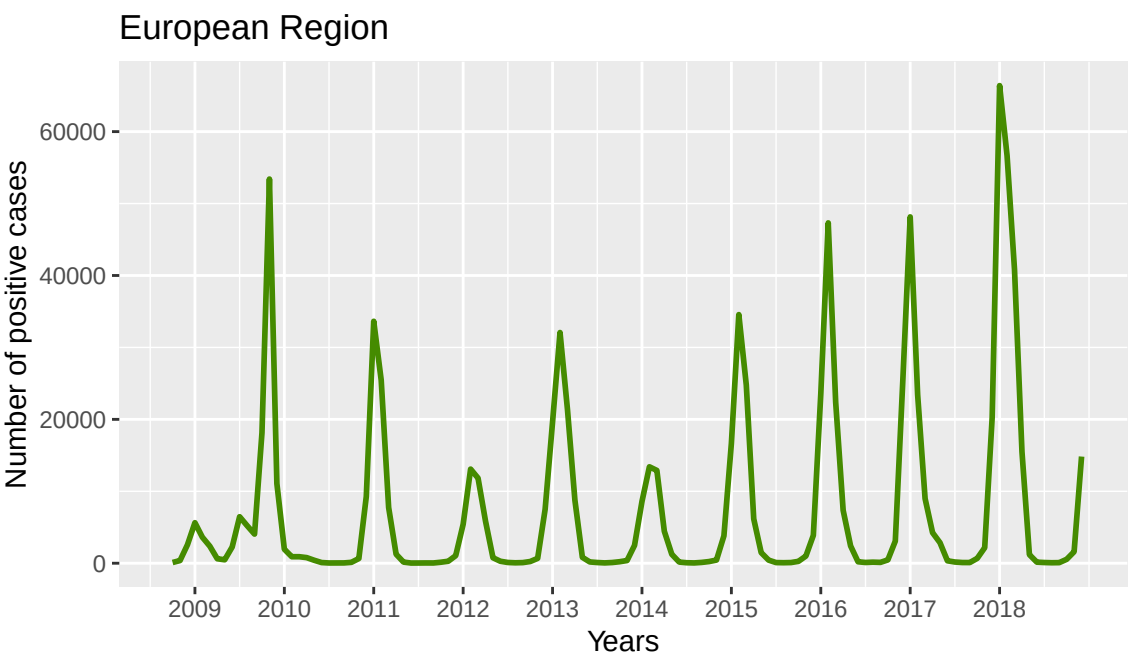

Comparing time series
Number of positive cases
1e+05
5e+04
0e+00
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
Years
Regions
NorthAmerica
Europe
SouthAmerica

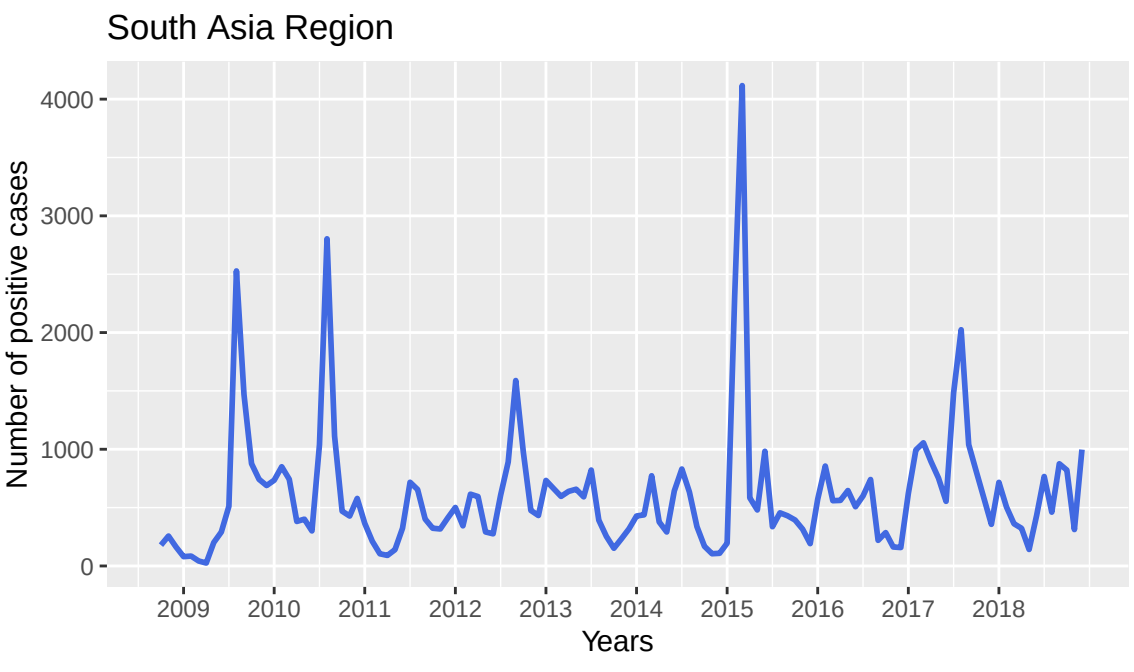

Brazil
Number of positive cases
2000
1500
1000
500
0
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
Years

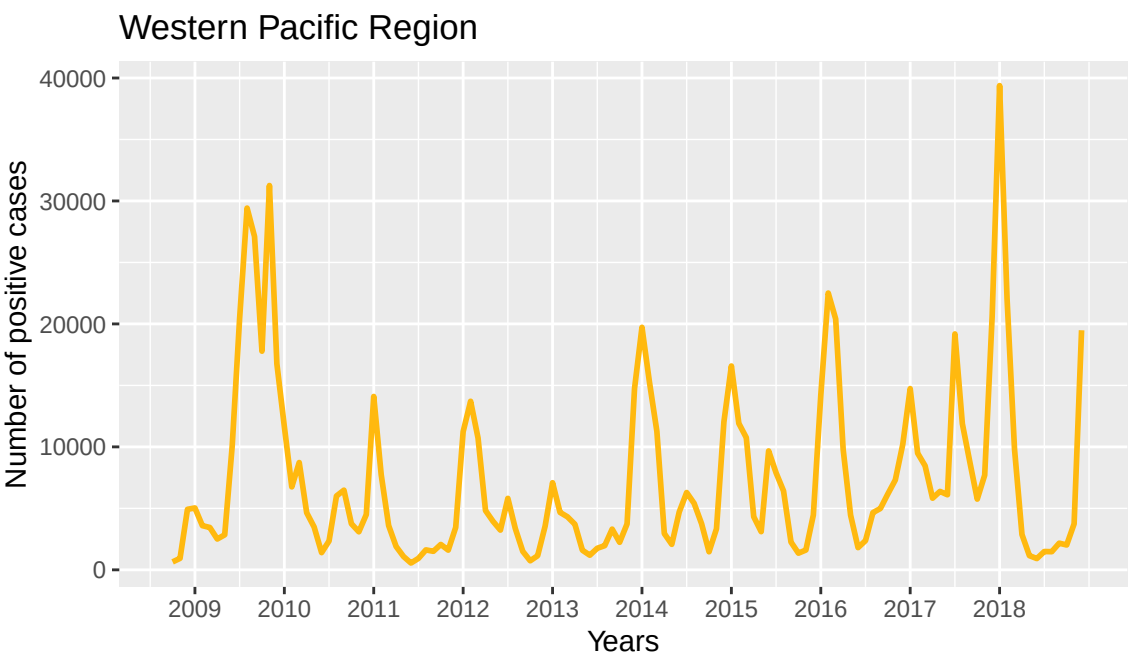

North America Region
Number of positive cases
1e+05
5e+04
0e+00
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
Years

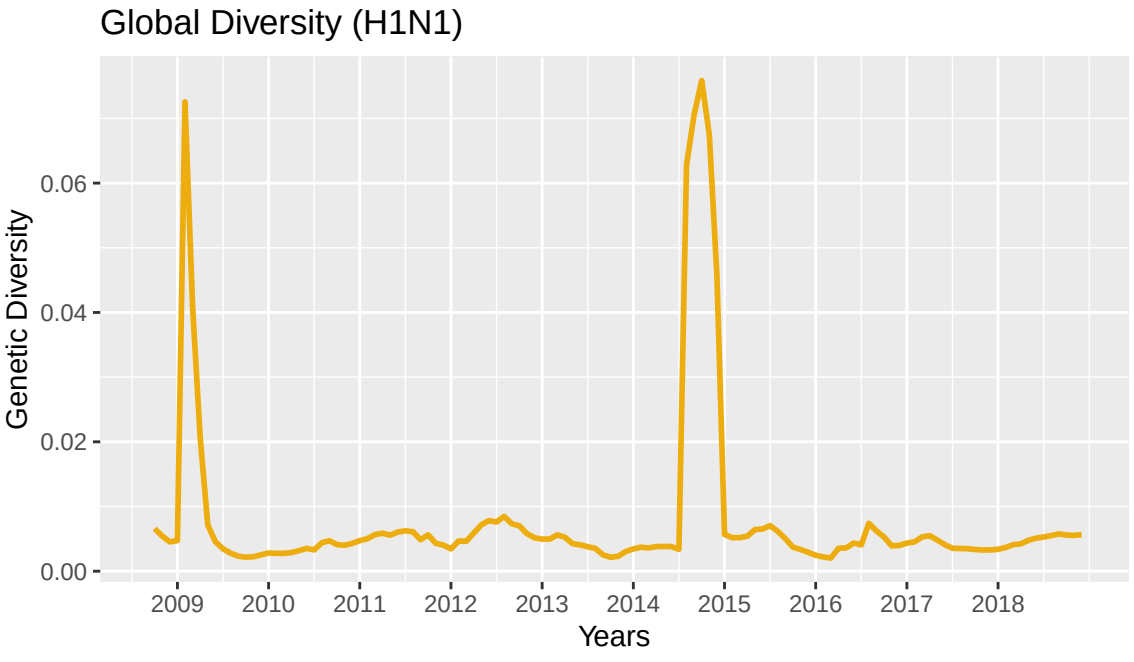

South America Region
Number of positive cases
9000
6000
3000
0
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
Years

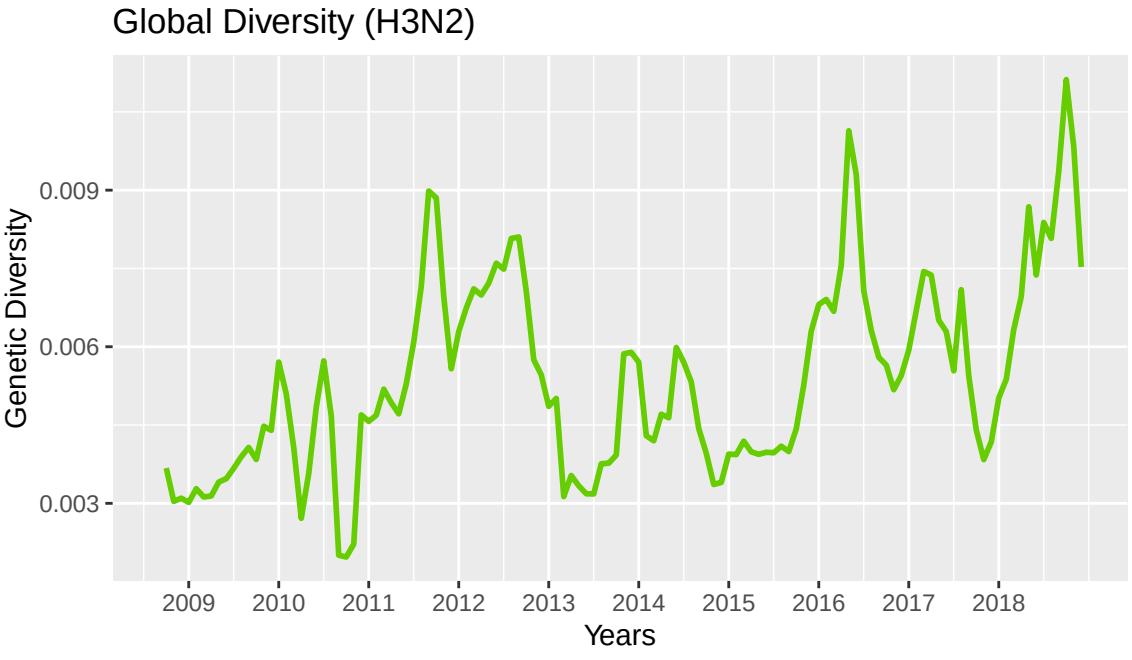

Central America Region
Number of positive cases
20000
15000
10000
5000
0
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
Years

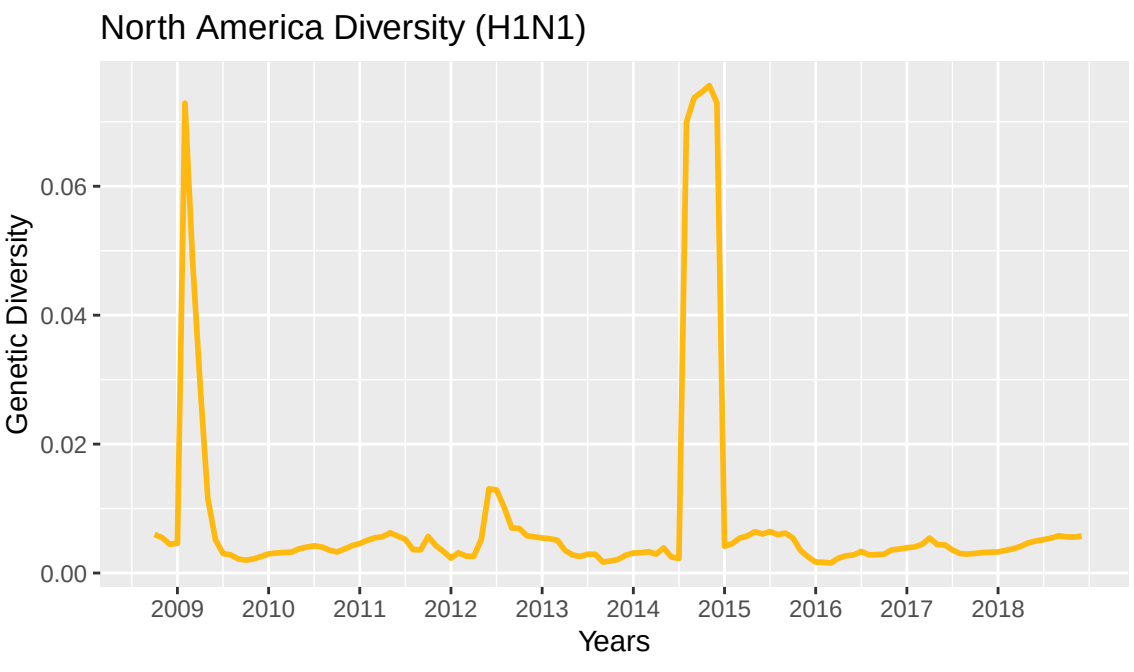

European Region
Number of positive cases
60000
40000
20000
0
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
Years

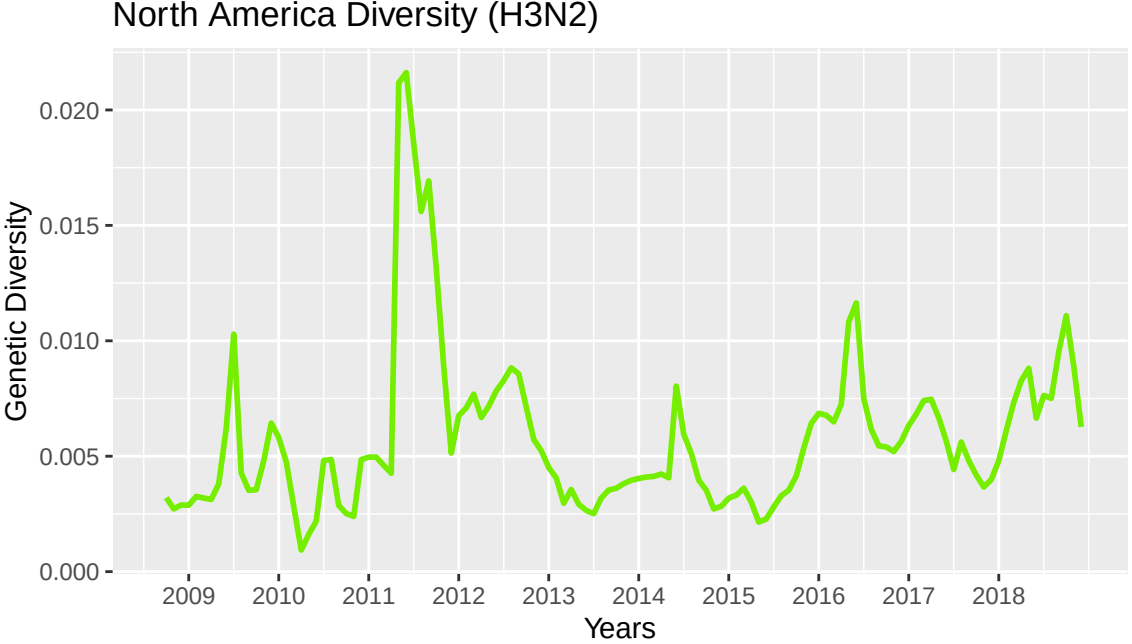

South Asia Region
Number of positive cases
4000
3000
2000
1000
0
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
Years

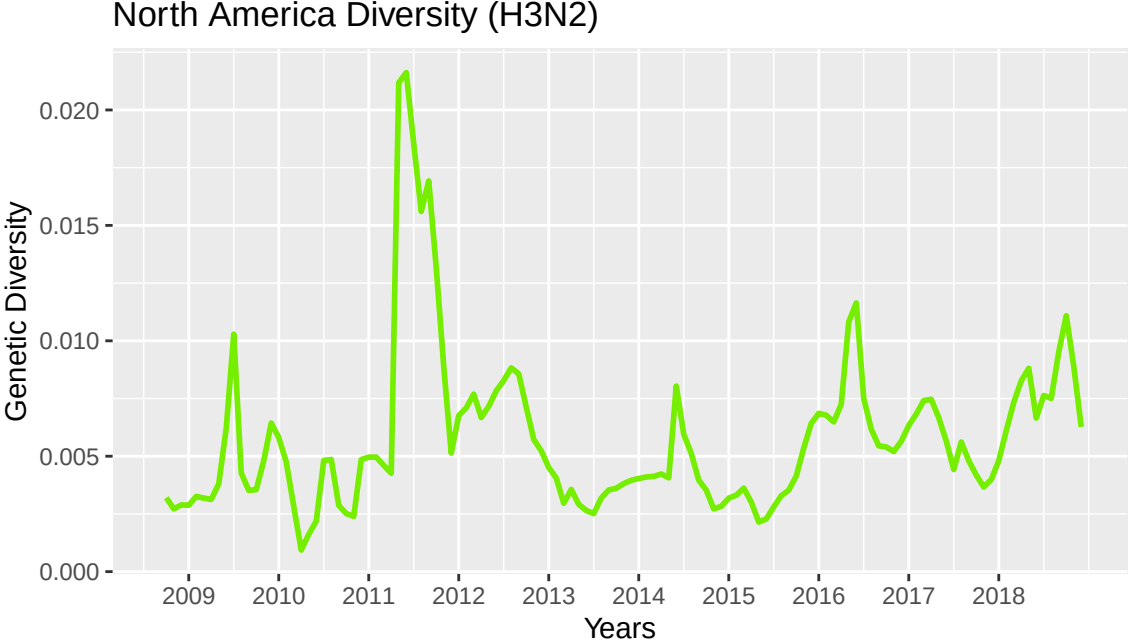

Western Pacific Region
Number of positive cases
40000
30000
20000
10000
0
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
Years

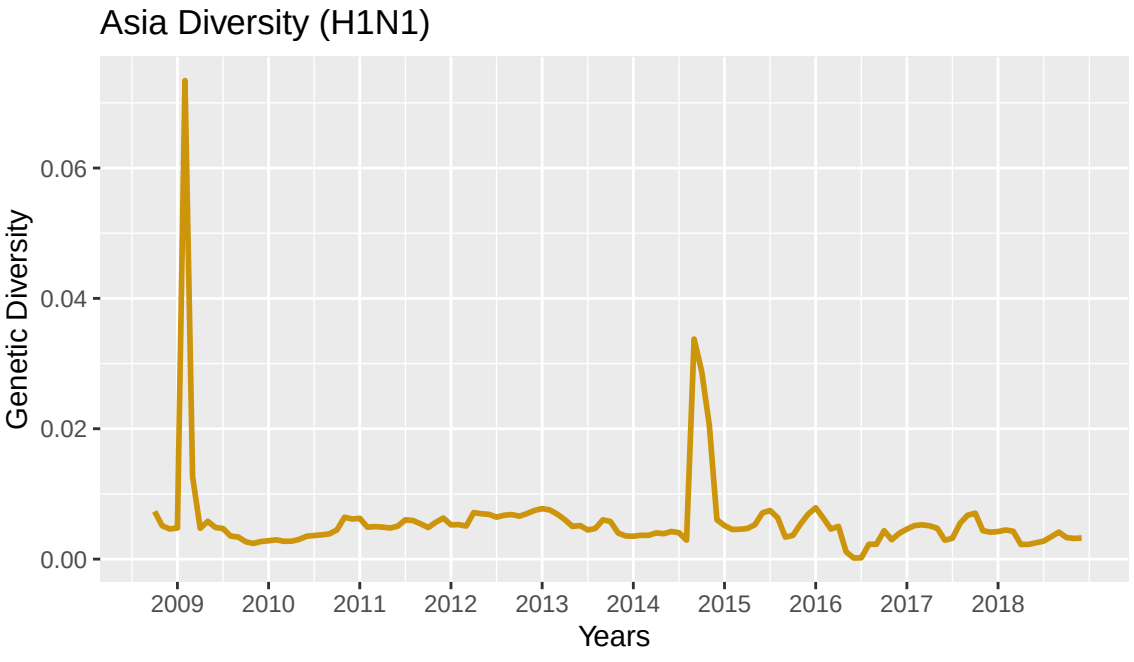

Global Diversity (H1N1)
Genetic Diversity
0.00
0.02
0.04
0.06
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
Years

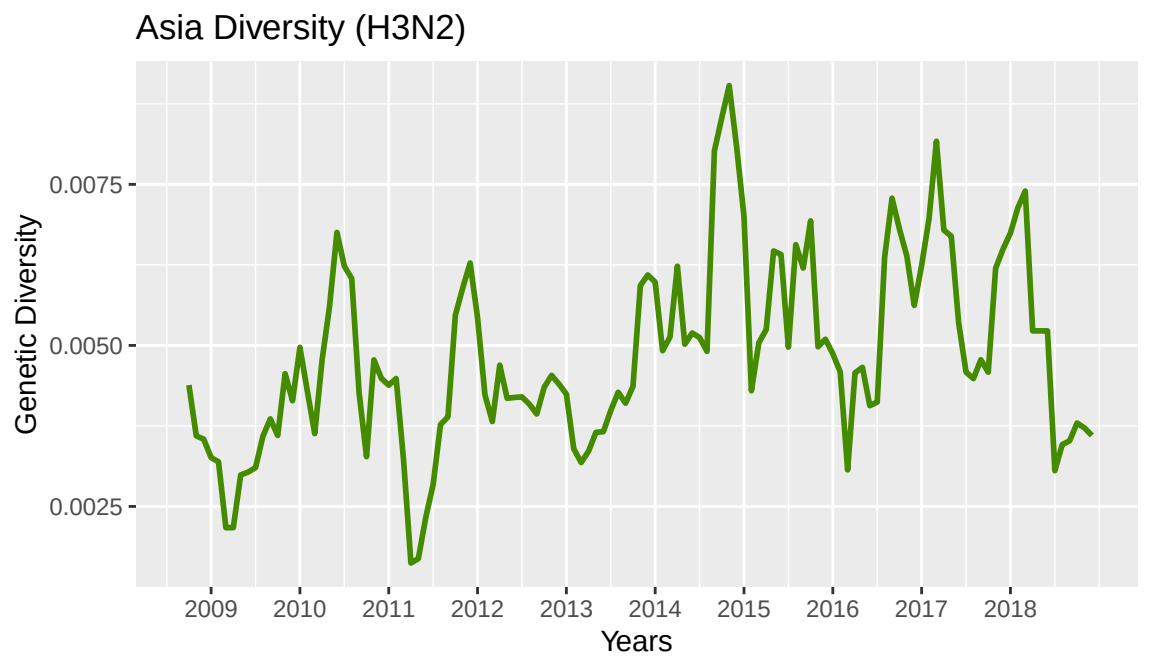

Global Diversity (H3N2)
Genetic Diversity
0.009
0.006
0.003
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
Years

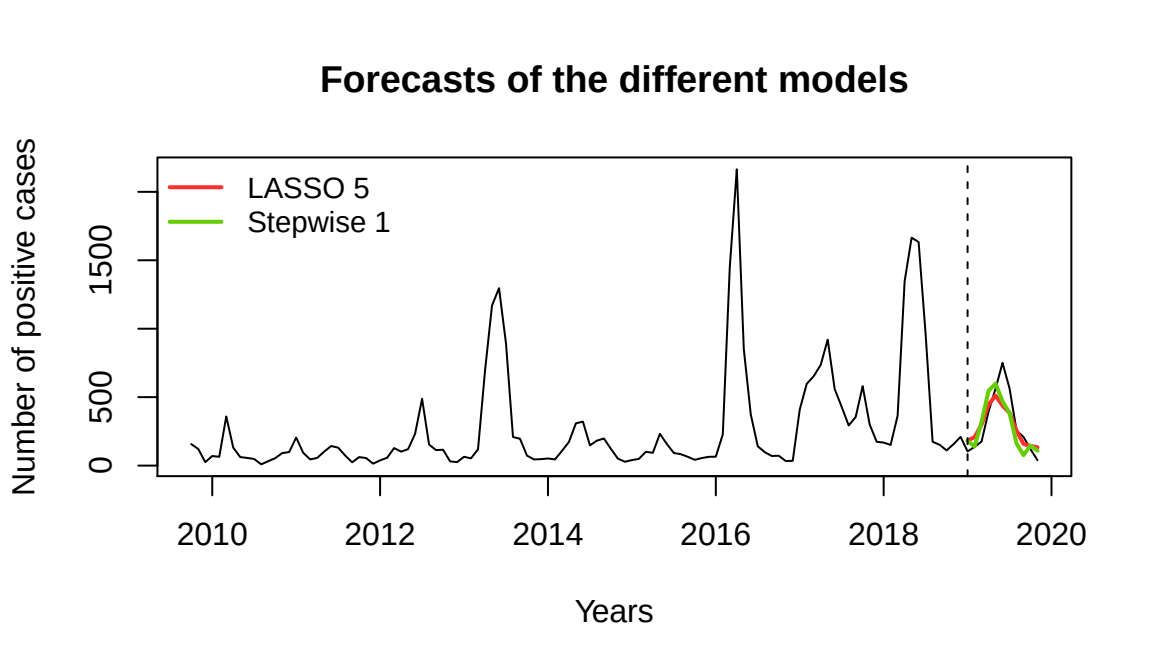

North America Diversity (H1N1)
Genetic Diversity
0.00
0.02
0.04
0.06
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
Years

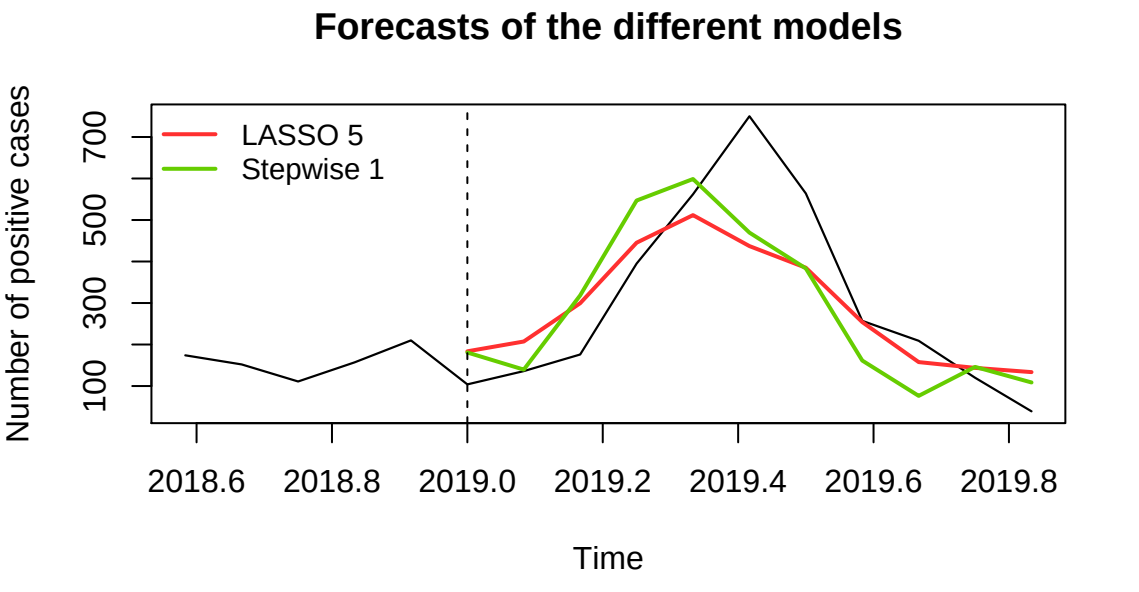

North America Diversity (H3N2)
Genetic Diversity
0.020
0.015
0.010
0.005
0.000
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
Years

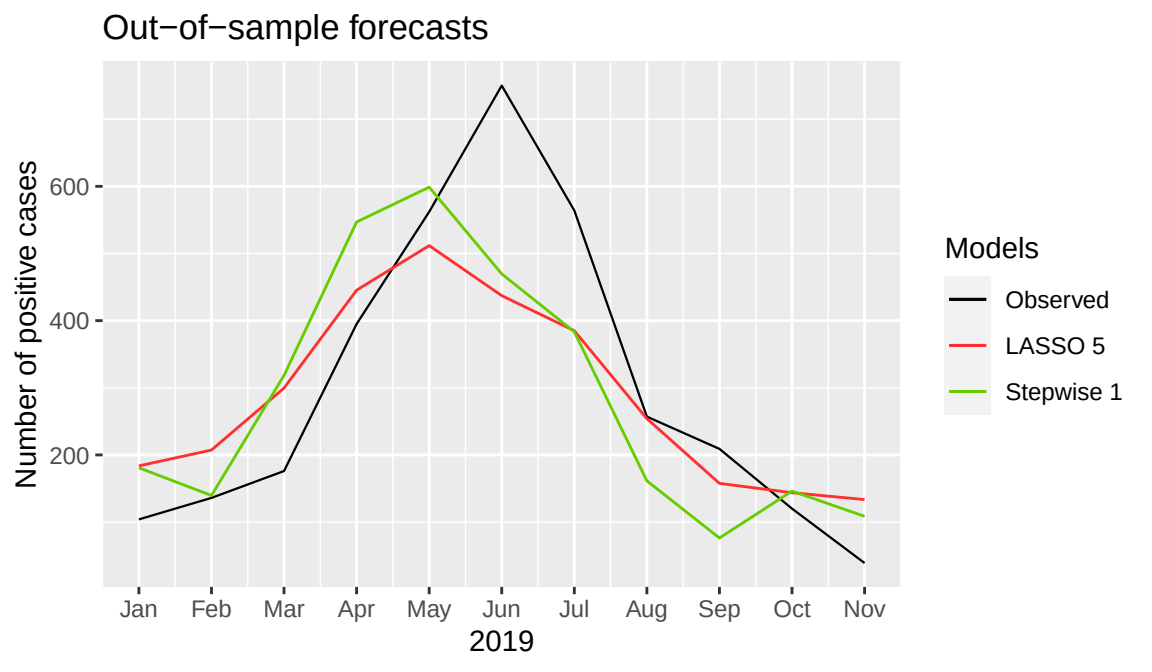

North America Diversity (H3N2)
Genetic Diversity
0.020
0.015
0.010
0.005
0.000
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
Years

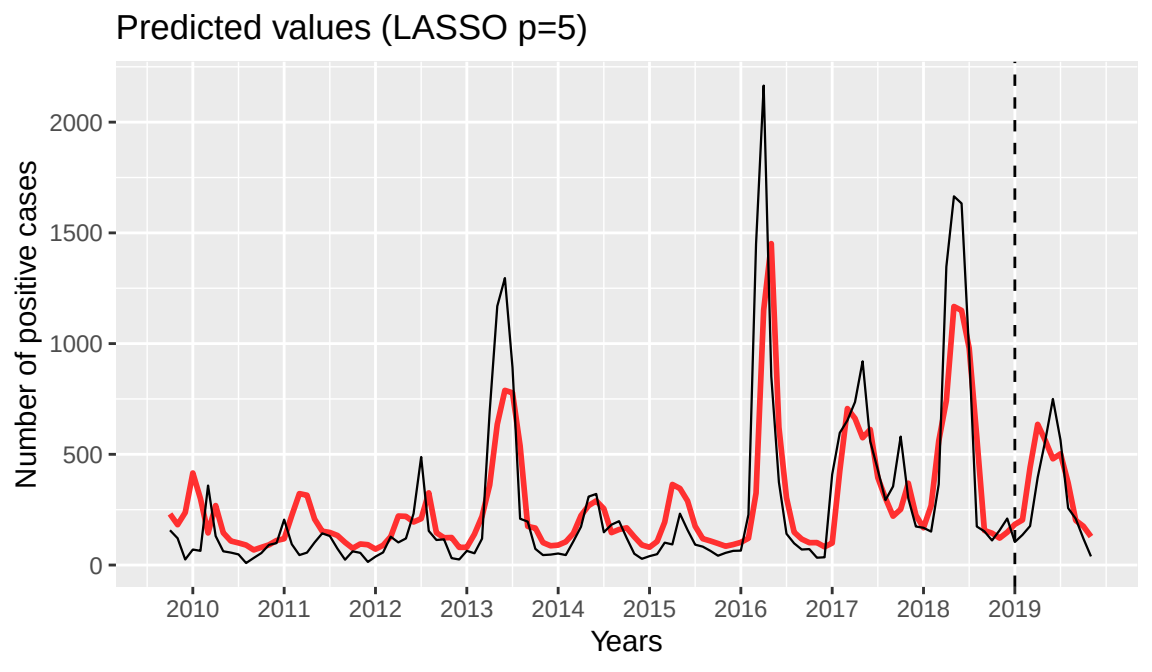

Asia Diversity (H1N1)
Genetic Diversity
0.06
0.04
0.02
0.00
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
Years

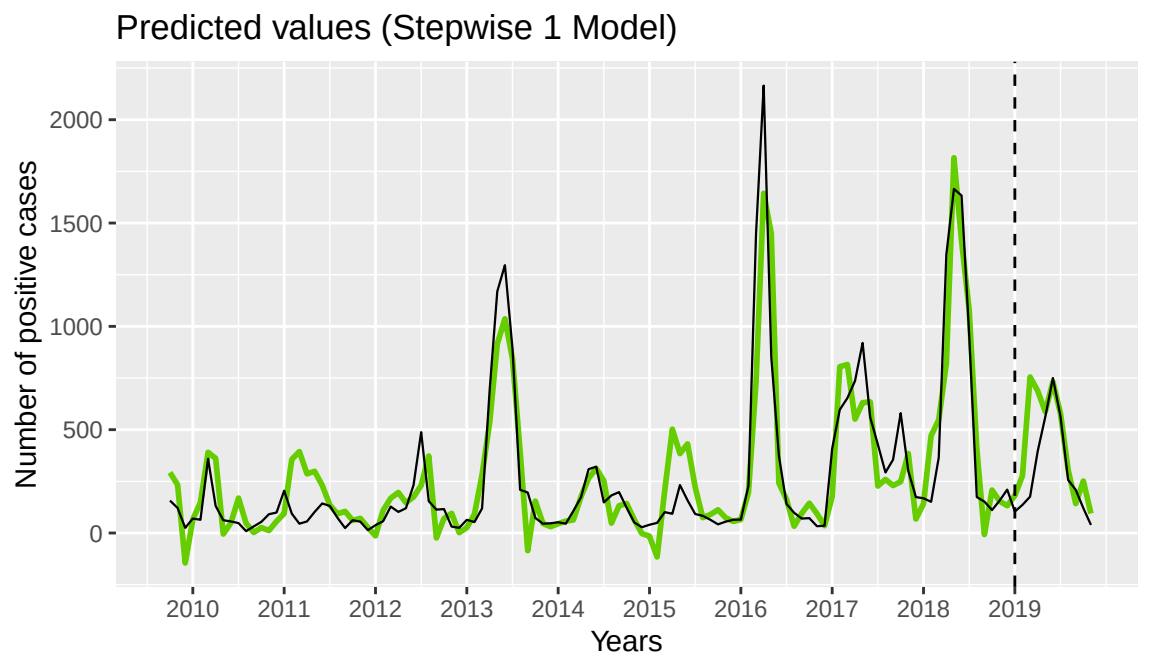

Asia Diversity (H3N2)
Genetic Diversity
0.0075
0.0050
0.0025
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
Years

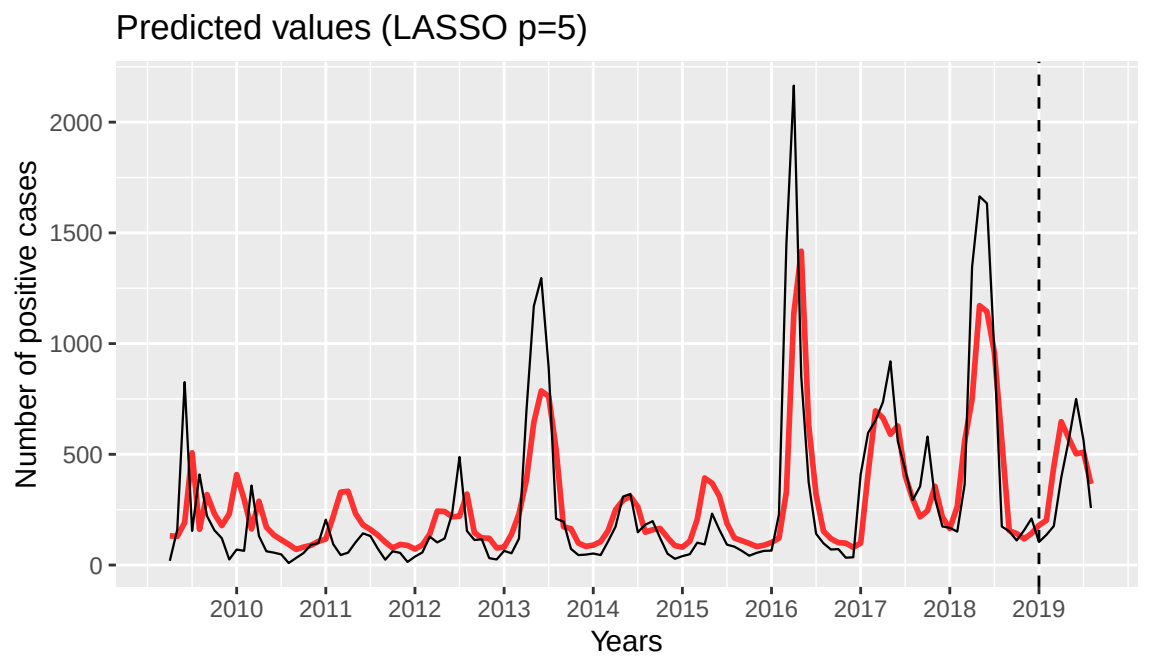

Forecasts of the different models
LASSO 5
Stepwise 1
Number of positive cases
0
500
1500
2010
2012
2014
2016
2018
2020
Years

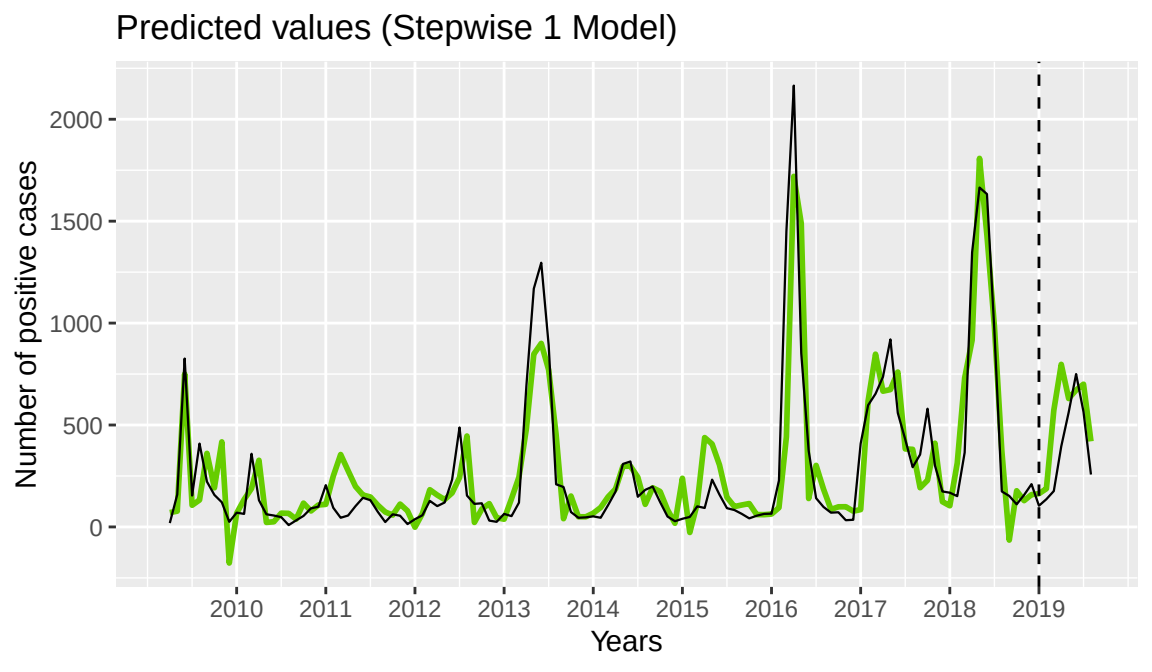

Forecasts of the different models
LASSO 5
Stepwise 1
Number of positive cases
100
300
500
700
2018.6
2018.8
2019.0
2019.2
2019.4
2019.6
2019.8
Time

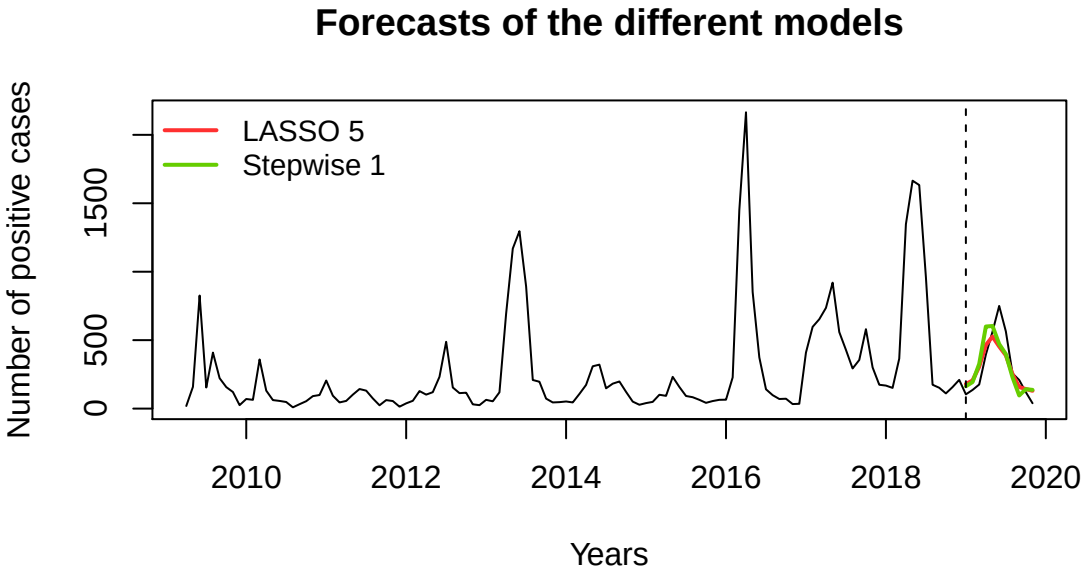

Out−of−sample forecasts
Number of positive cases
600
400
200
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
2019
Models
Observed
LASSO 5
Stepwise 1

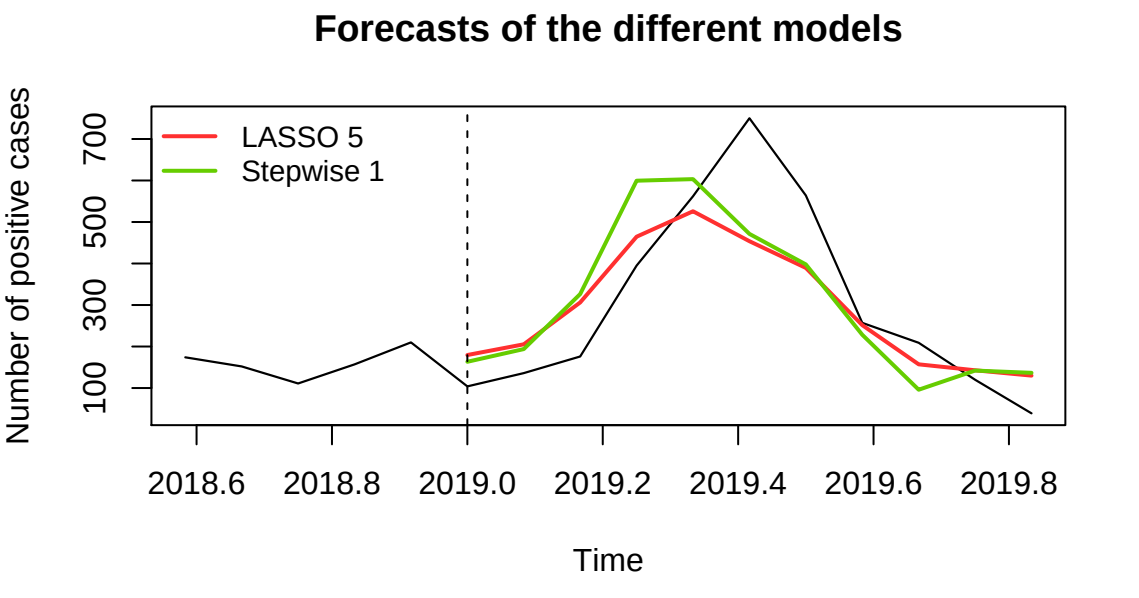

Predicted values (LASSO p=5)
Number of positive cases
2000
1500
1000
500
0
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
Years

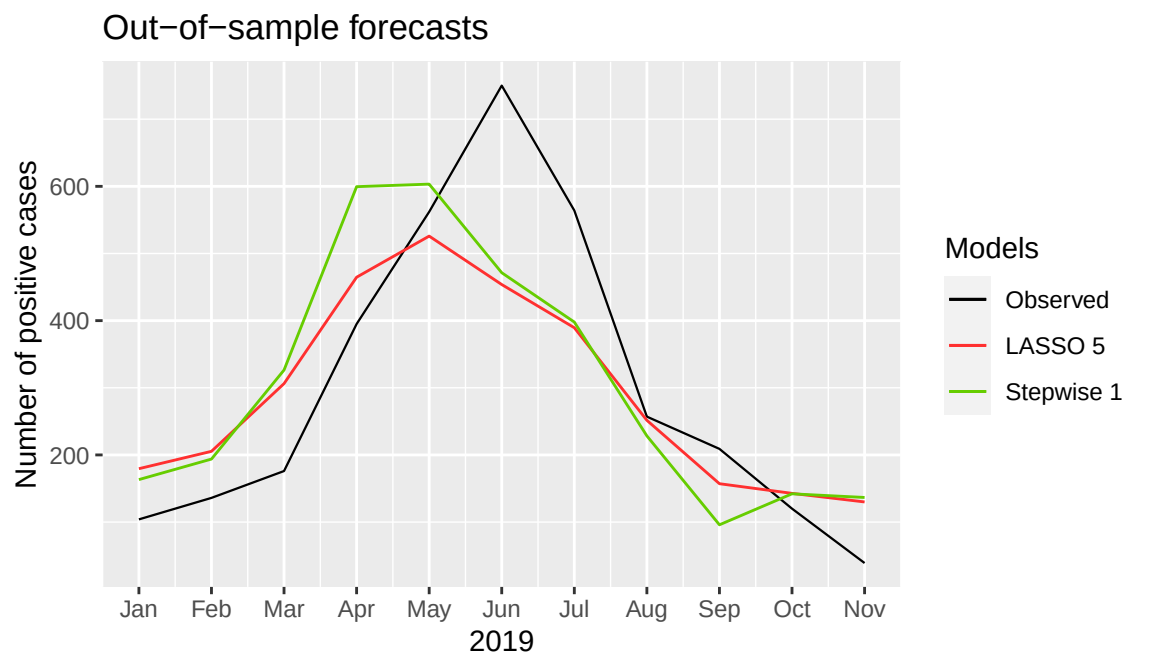

Predicted values (Stepwise 1 Model)
Number of positive cases
2000
1500
1000
500
0
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
Years

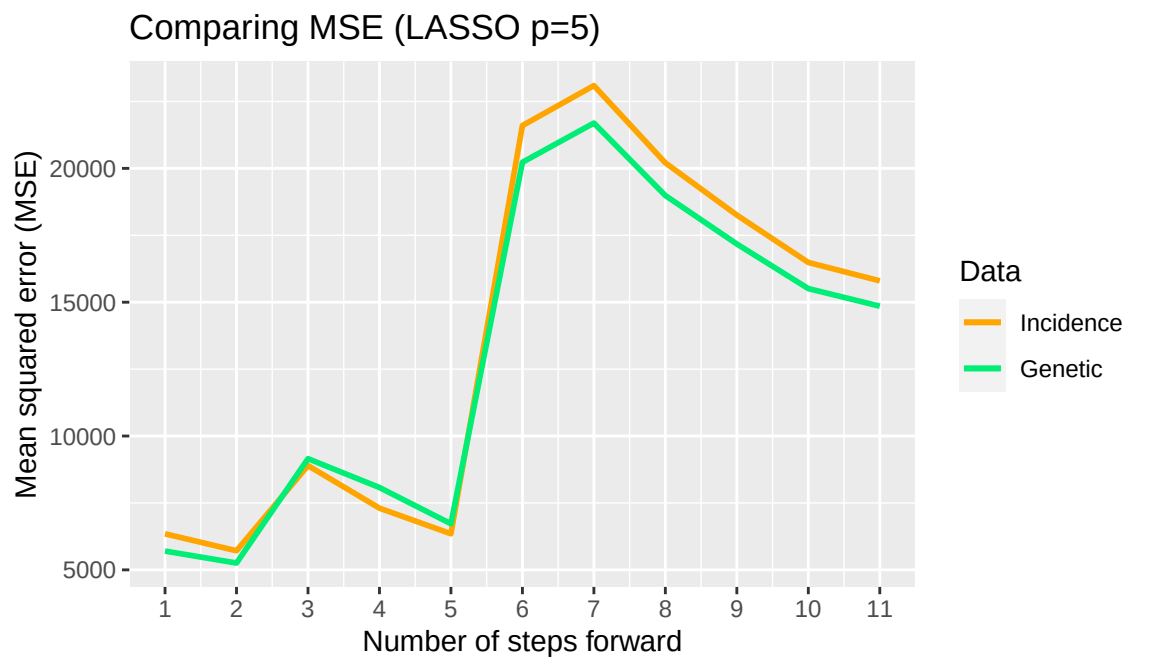

Predicted values (LASSO p=5)
Number of positive cases
2000
1500
1000
500
0
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
Years

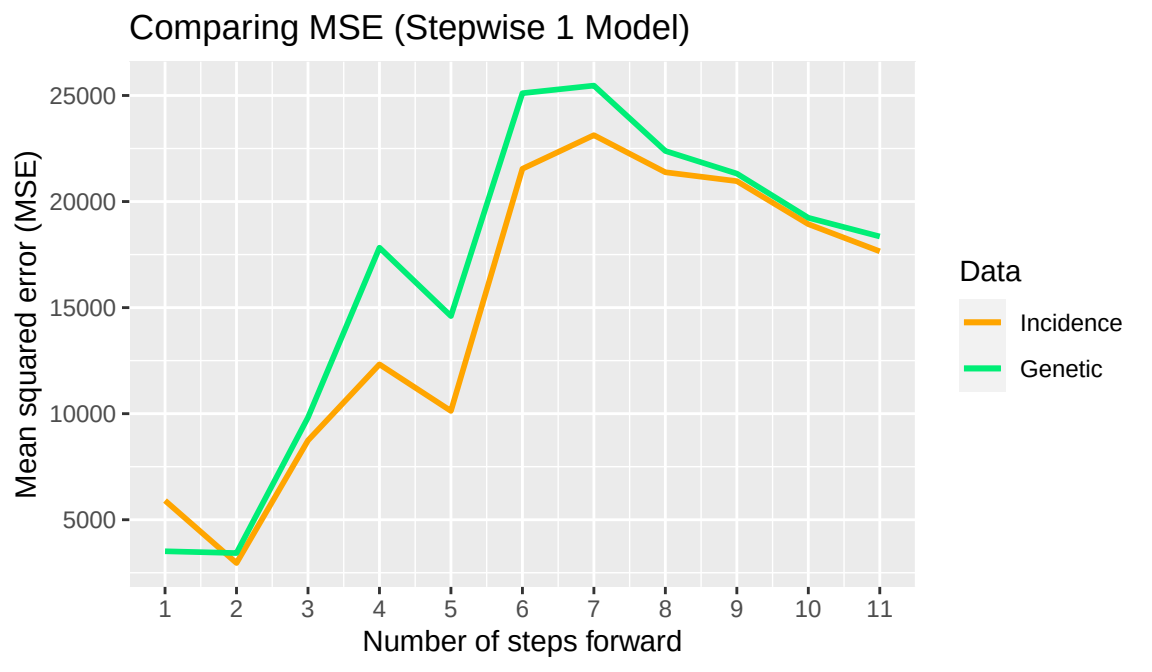

Predicted values (Stepwise 1 Model)
Number of positive cases
2000
1500
1000
500
0
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
Years

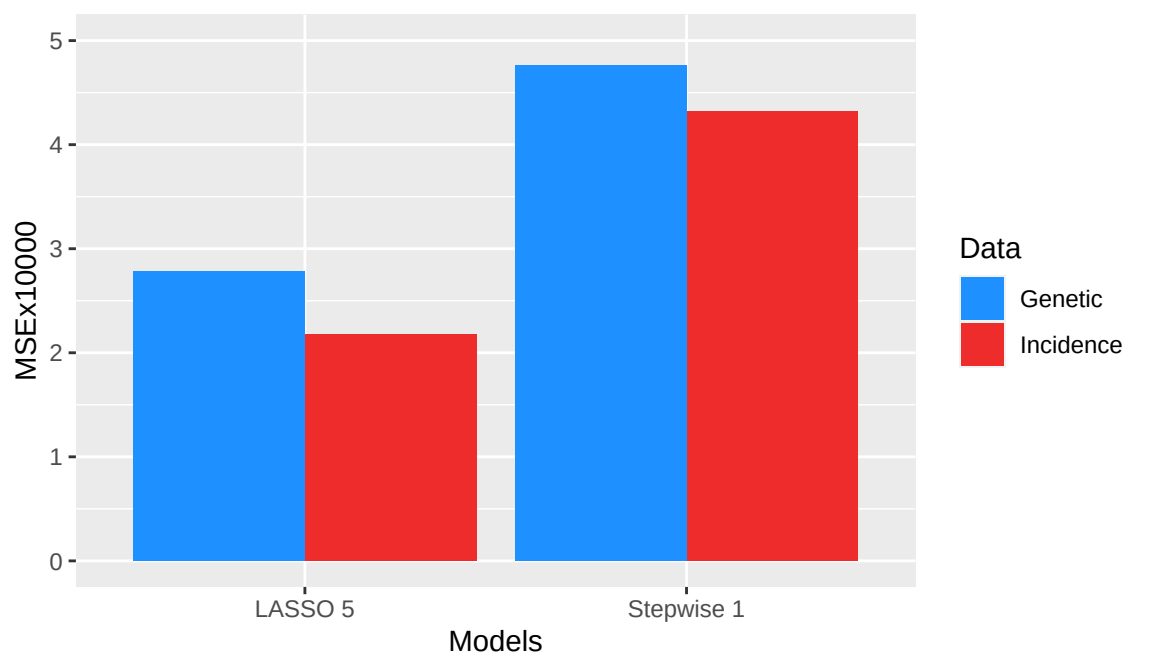

Forecasts of the different models
LASSO 5
Stepwise 1
Number of positive cases
0
500
1500
2010
2012
2014
2016
2018
2020
Years

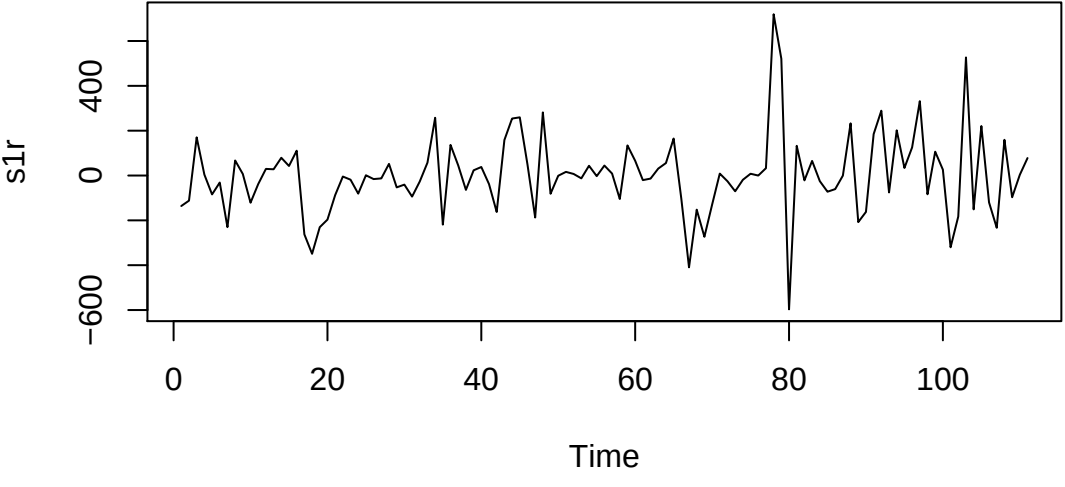

Forecasts of the different models
LASSO 5
Stepwise 1
Number of positive cases
100
300
500
700
2018.6
2018.8
2019.0
2019.2
2019.4
2019.6
2019.8
Time

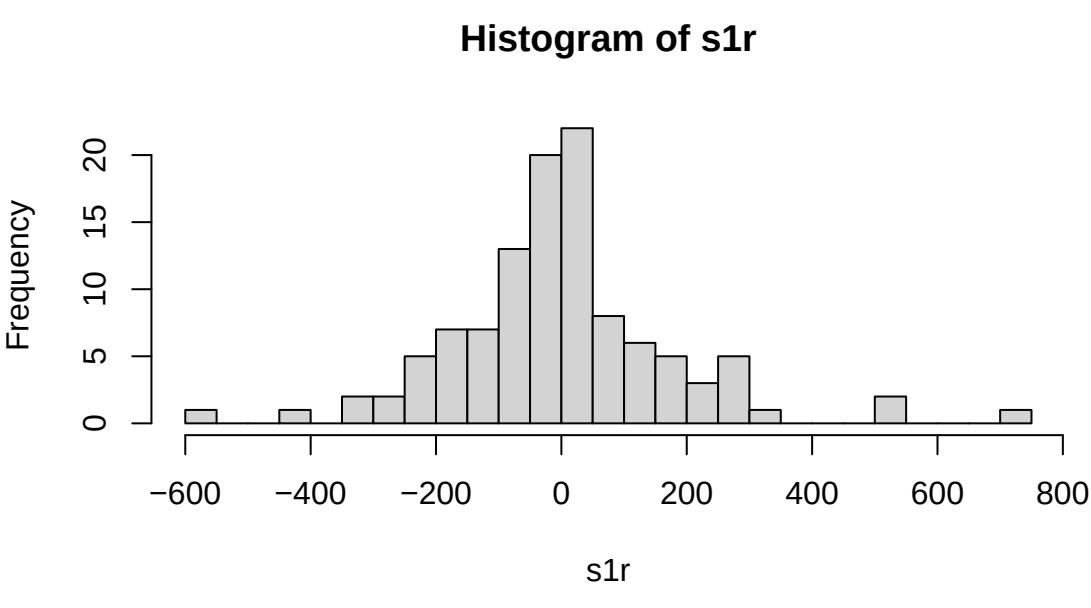

Out−of−sample forecasts
Number of positive cases
600
400
200
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
2019
Models
Observed
LASSO 5
Stepwise 1

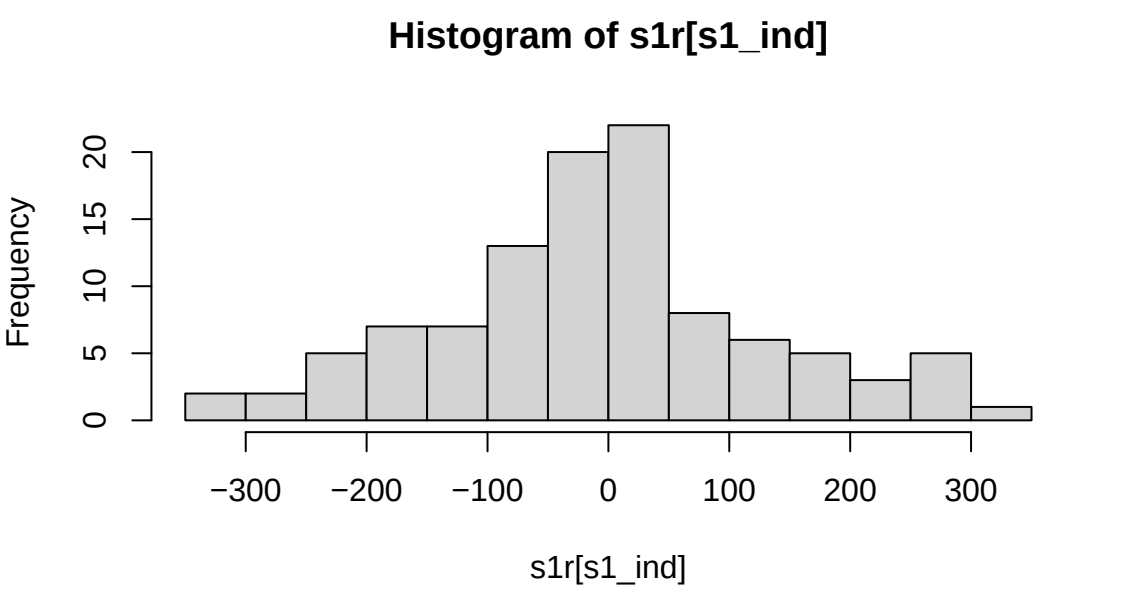

Comparing MSE (LASSO p=5)
Mean squared error (MSE)
20000
15000
10000
5000
1
2
3
4
5
6
7
8
9
10
11
Number of steps forward
Data
Incidence
Genetic

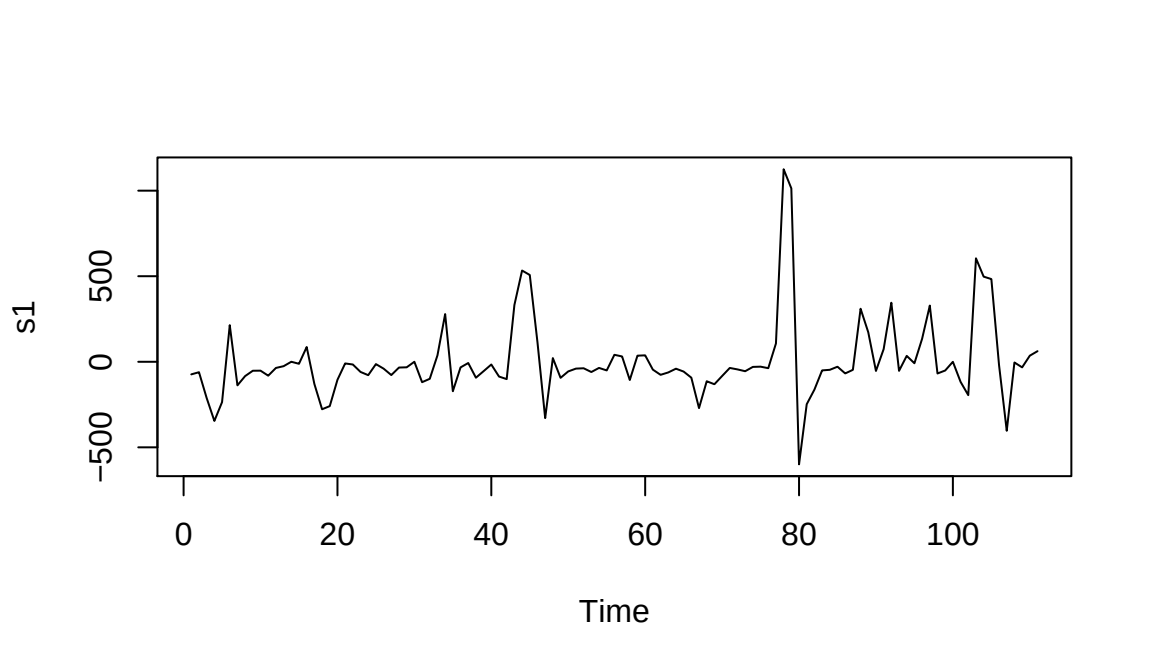

Comparing MSE (Stepwise 1 Model)
Mean squared error (MSE)
25000
20000
15000
10000
5000
1
2
3
4
5
6
7
8
9
10
11
Number of steps forward
Data
Incidence
Genetic

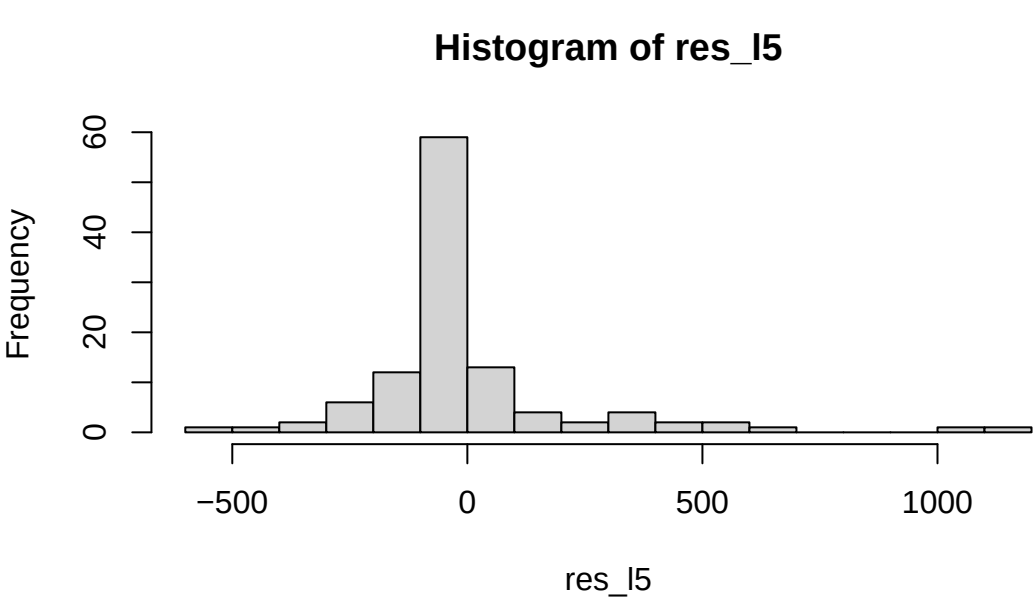

5
4
3
2
1
0
MSEx10000
LASSO 5
Stepwise 1
Models
Data
Genetic
Incidence

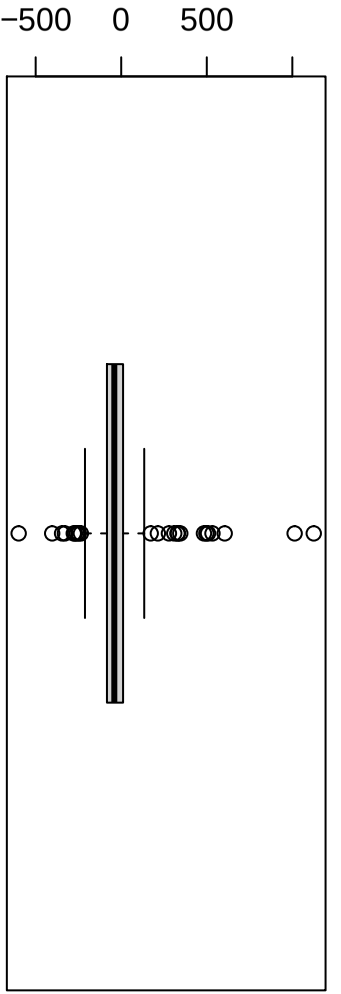

s1r
400
0
−600
0
20
40
60
80
100
Time

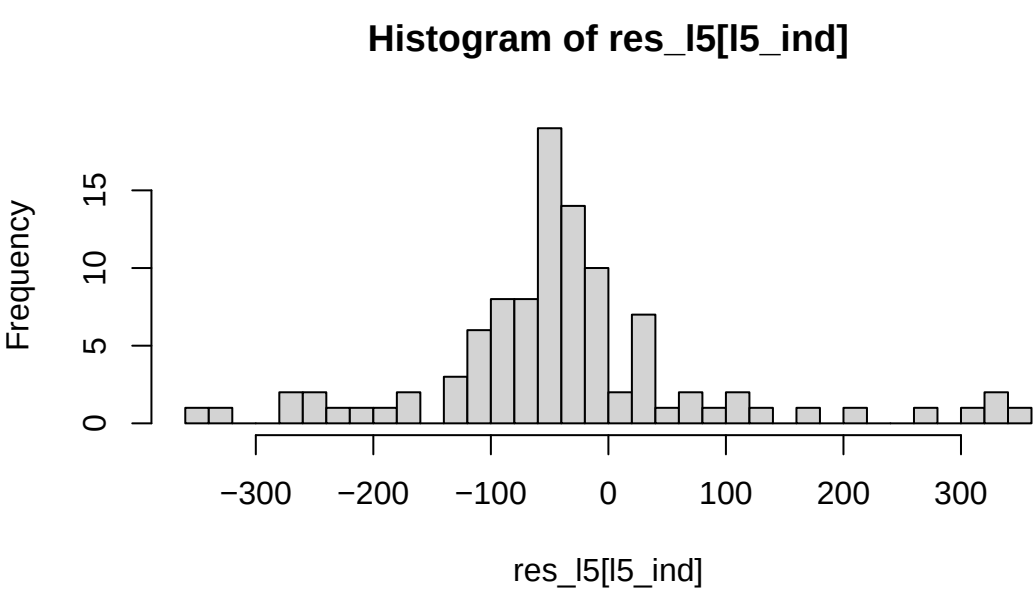

Histogram of s1r
Frequency
0
5
10
15
20
−600
−400
−200
0
200
400
600
800
s1r

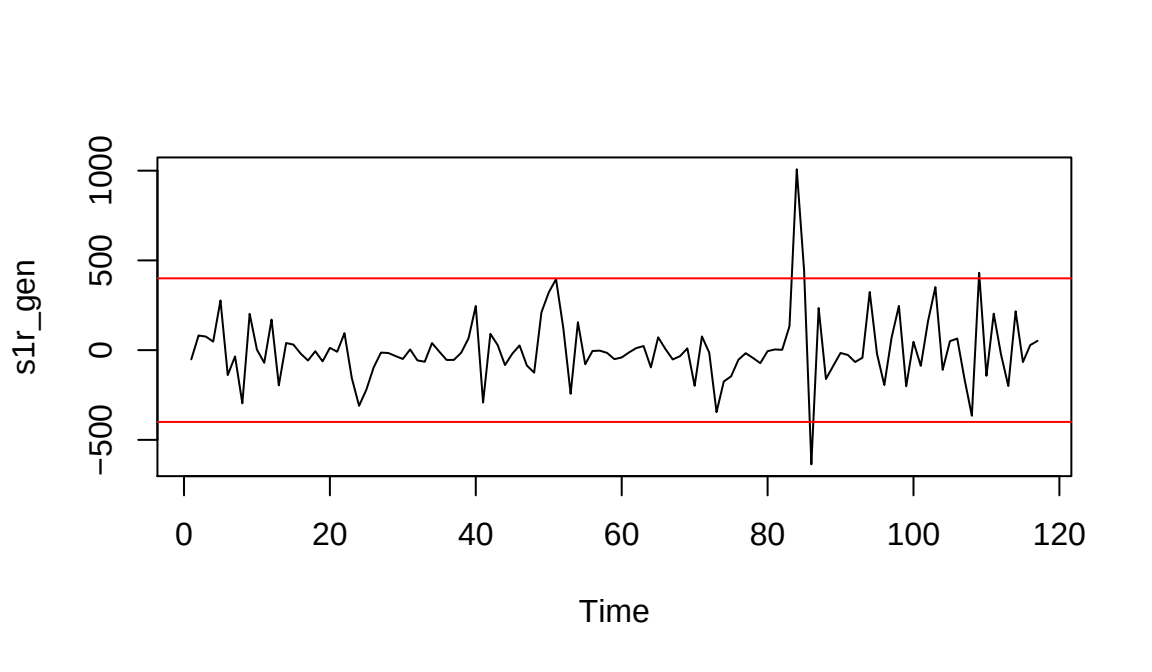

Histogram of s1r[s1_ind]
Frequency
0
5
10
15
20
−300
−200
−100
0
100
200
300
s1r[s1_ind]

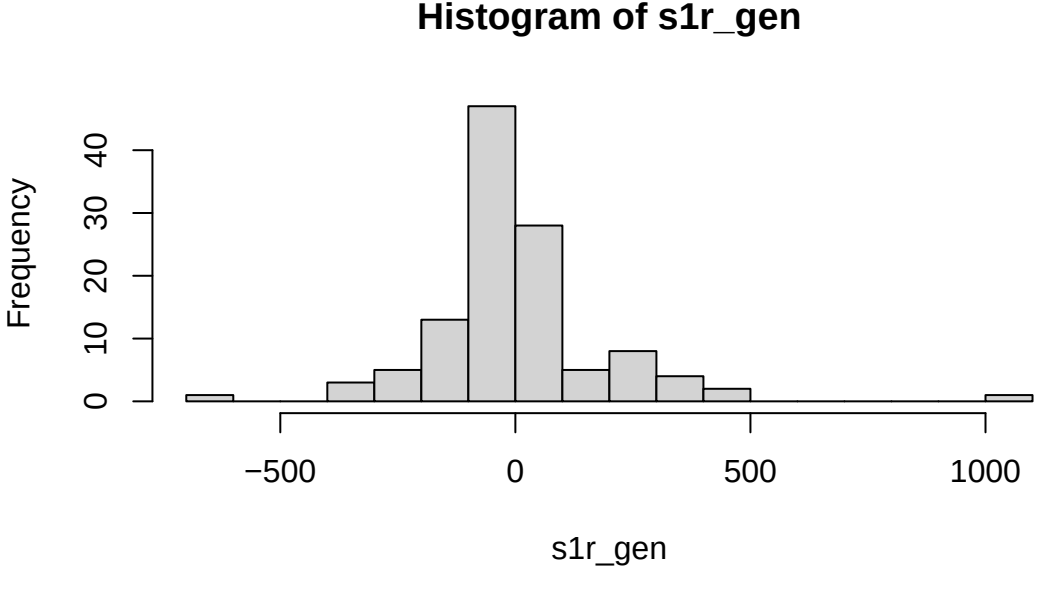

s1
−500
0
500
0
20
40
60
80
100
Time

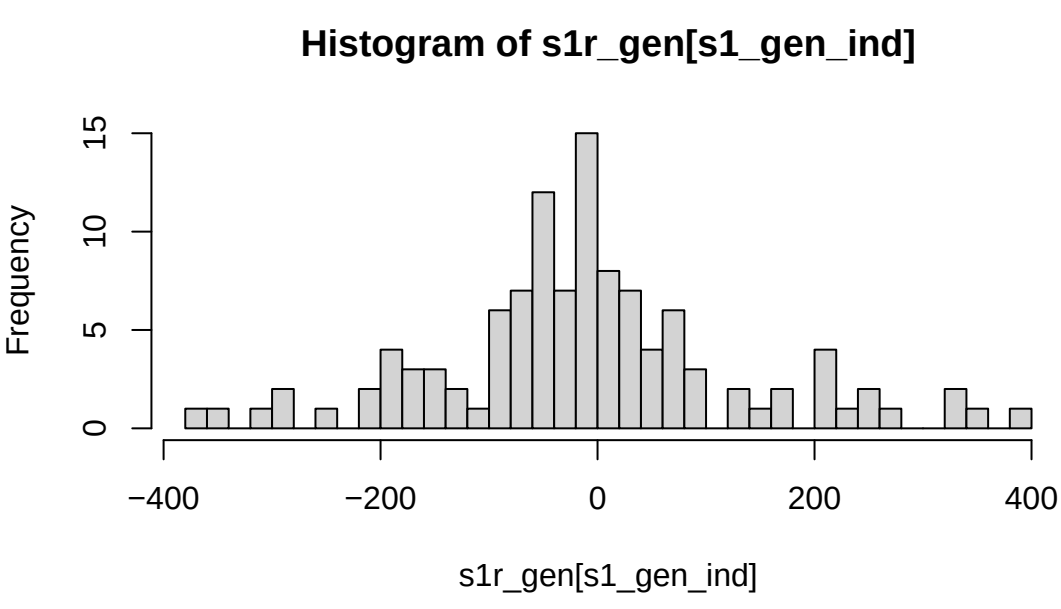

Histogram of res_l5
Frequency
0
20
40
60
−500
0
500
1000
res_l5

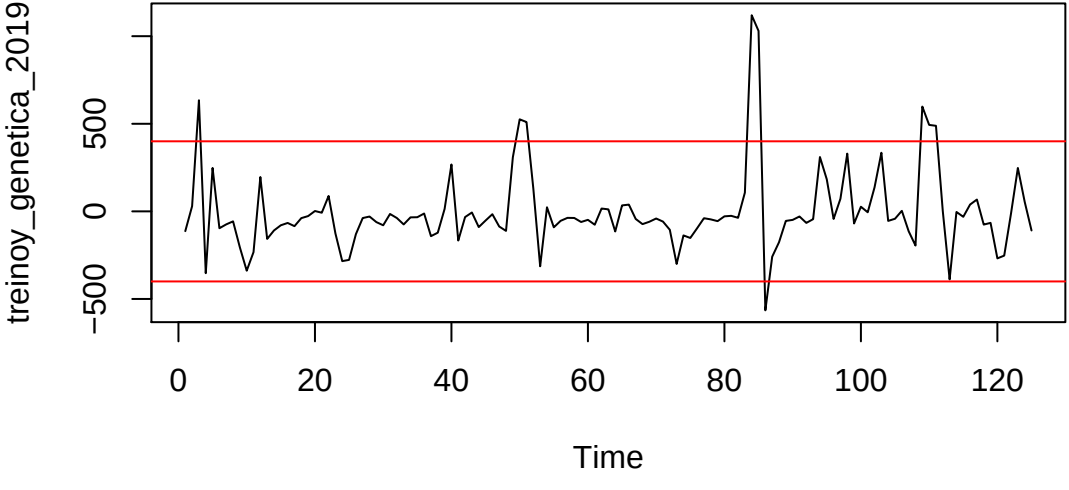

−500
0
500

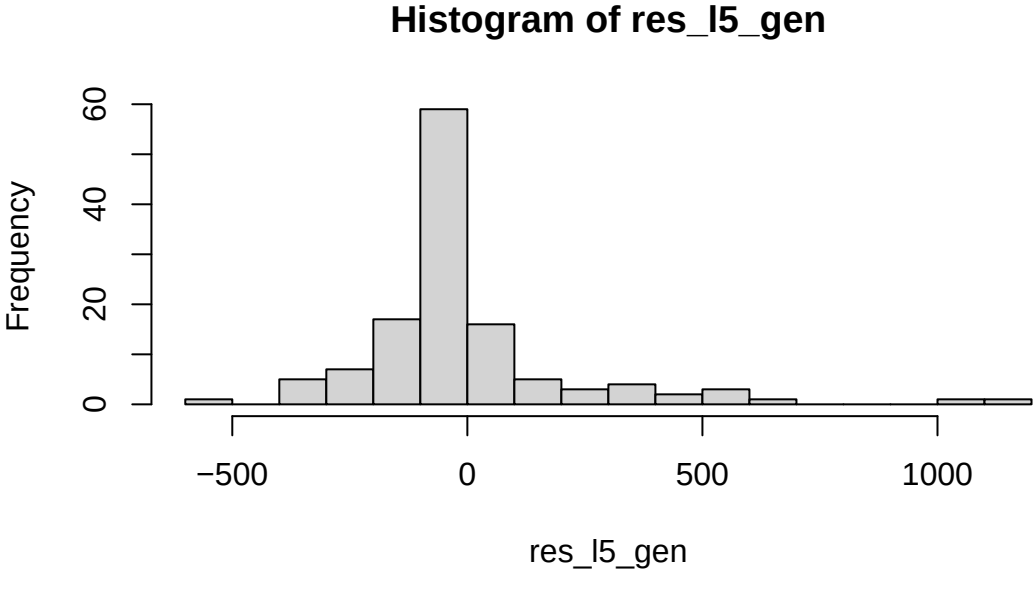

Histogram of res_l5[l5_ind]
Frequency
0
5
10
15
−300
−200
−100
0
100
200
300
res_l5[l5_ind]

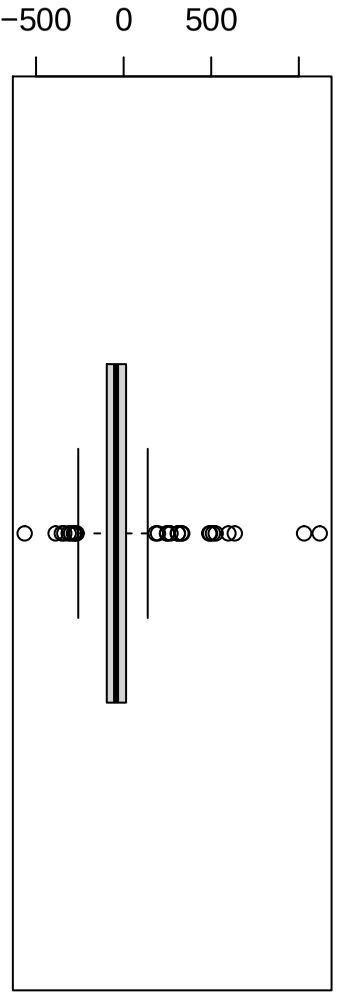

s1r_gen
1000
500
0
−500
0
20
40
60
80
100
120
Time

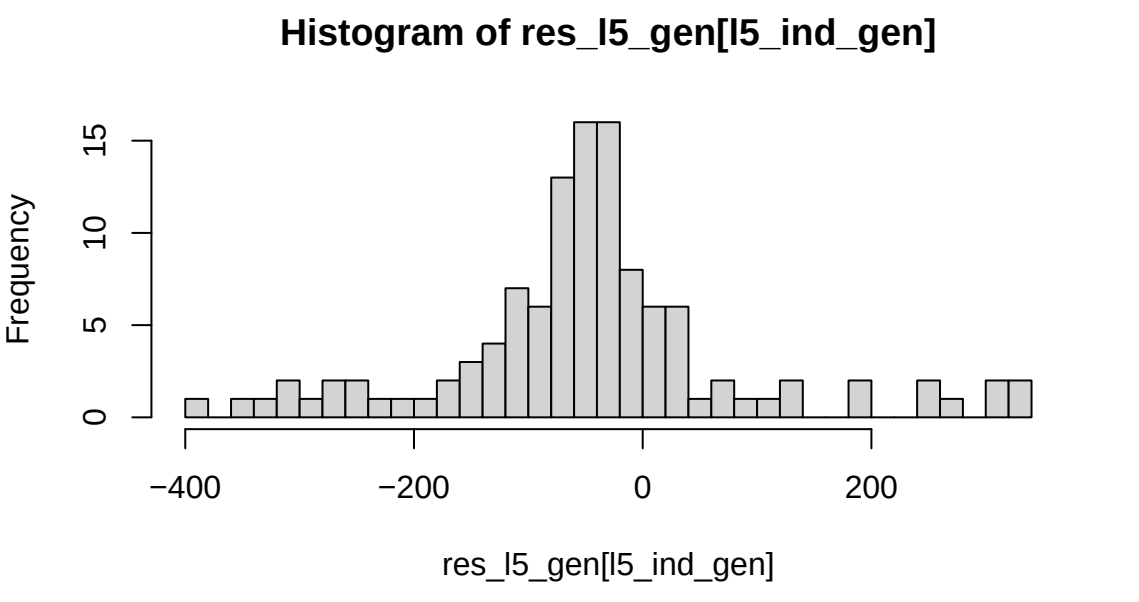

Histogram of s1r_gen
Frequency
0
10
20
30
40
−500
0
500
1000
s1r_gen

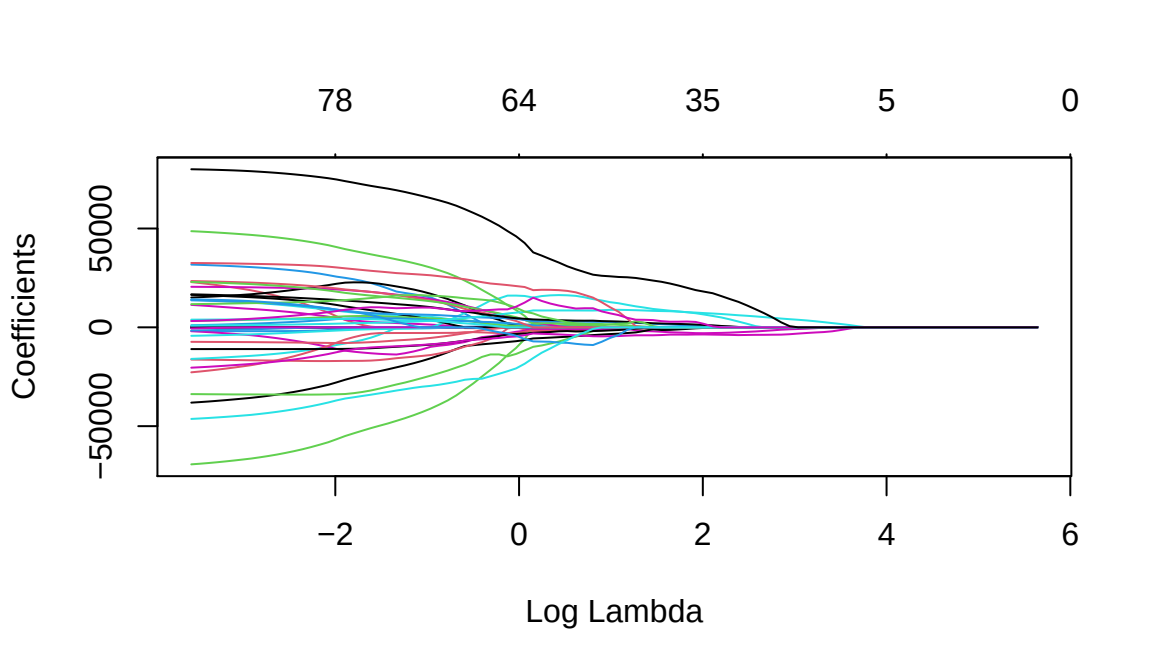

Histogram of s1r_gen[s1_gen_ind]
Frequency
0
5
10
15
−400
−200
0
200
400
s1r_gen[s1_gen_ind]

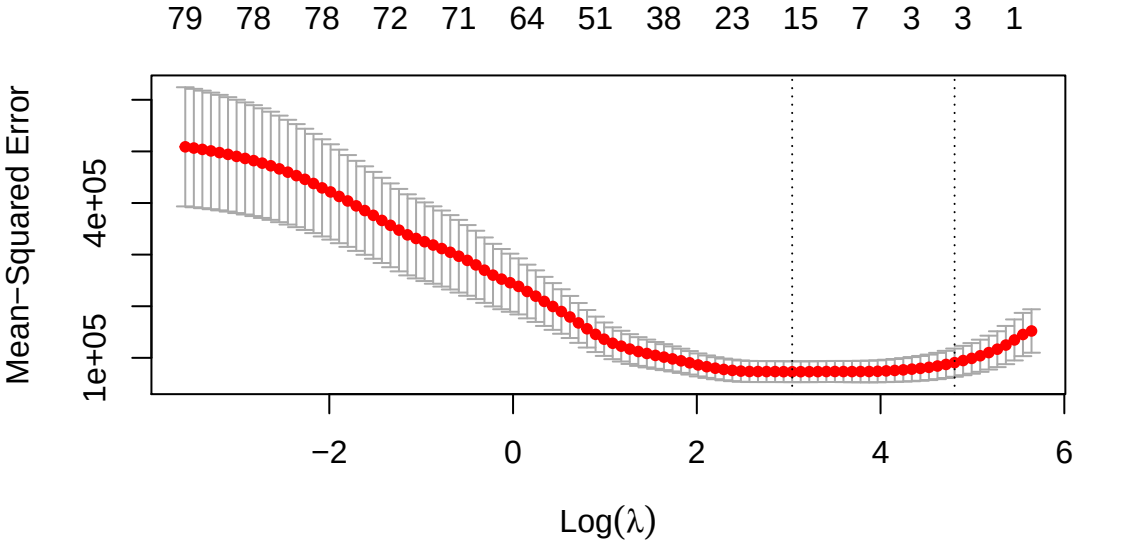

treinoy_genetica_2019
−500
0
500
0
20
40
60
80
100
120
Time

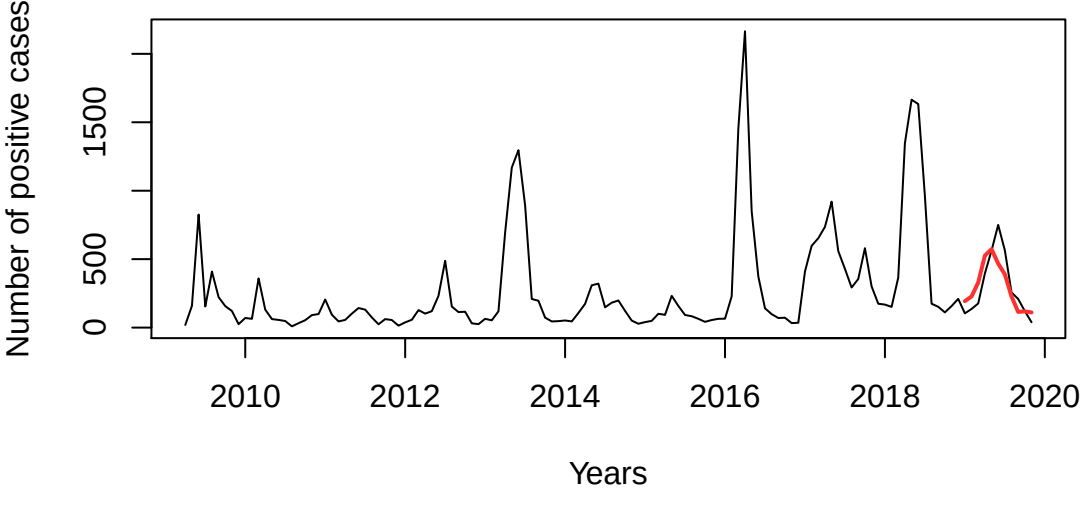

Histogram of res_l5_gen
Frequency
0
20
40
60
−500
0
500
1000
res_l5_gen

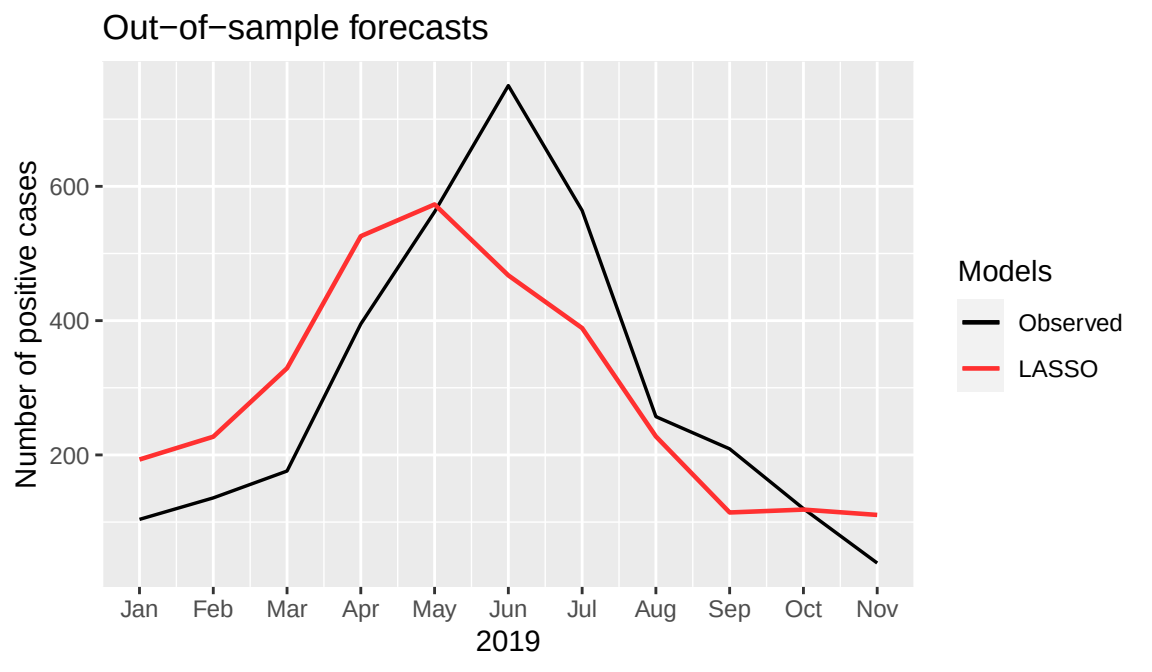

−500
0
500

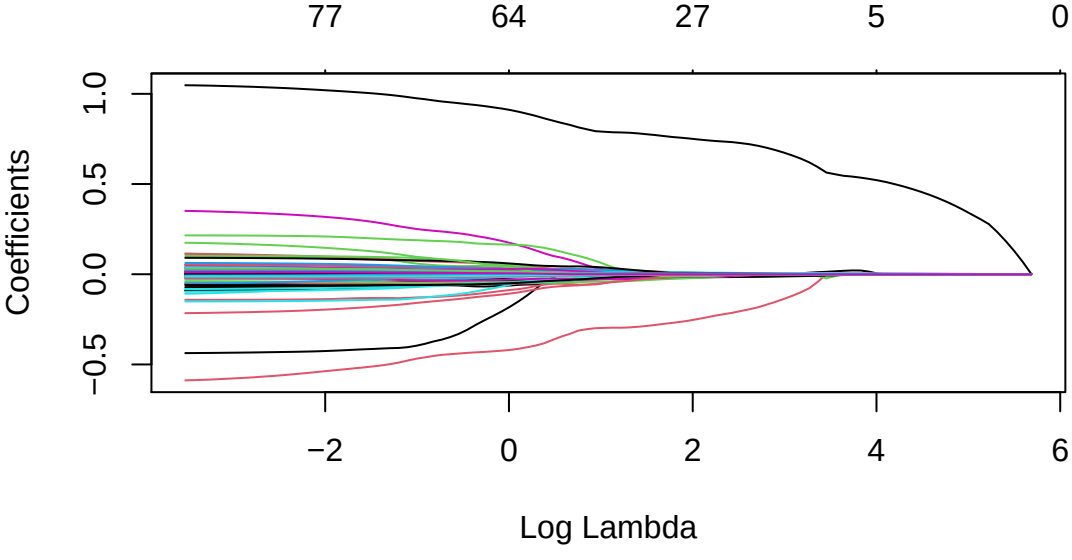

Histogram of res_l5_gen[l5_ind_gen]
Frequency
0
5
10
15
−400
−200
0
200
res_l5_gen[l5_ind_gen]

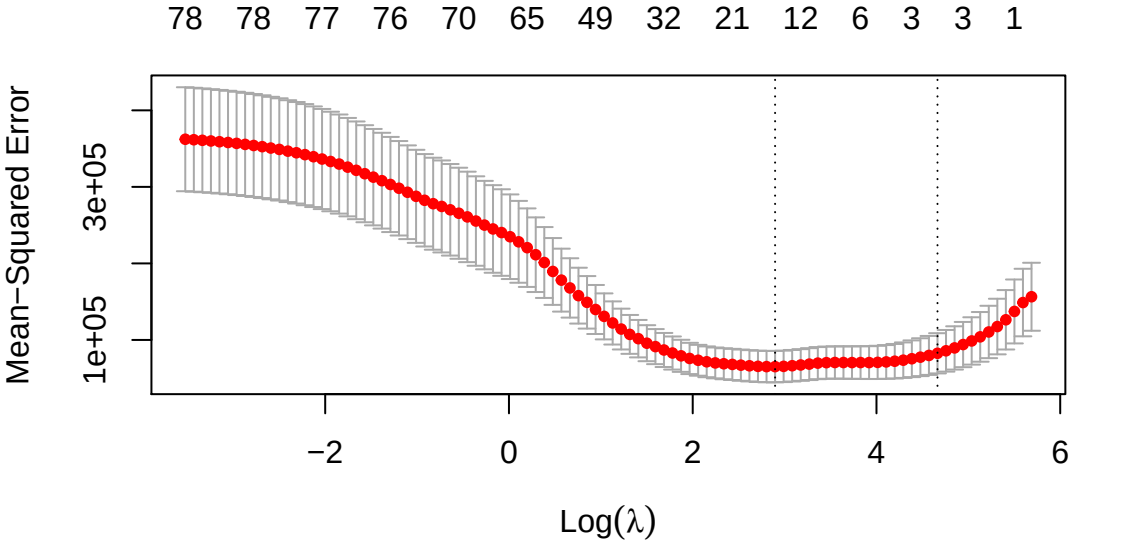

78
64
35
5
0
Coefficients
50000
0
−50000
−2
0
2
4
6
Log Lambda

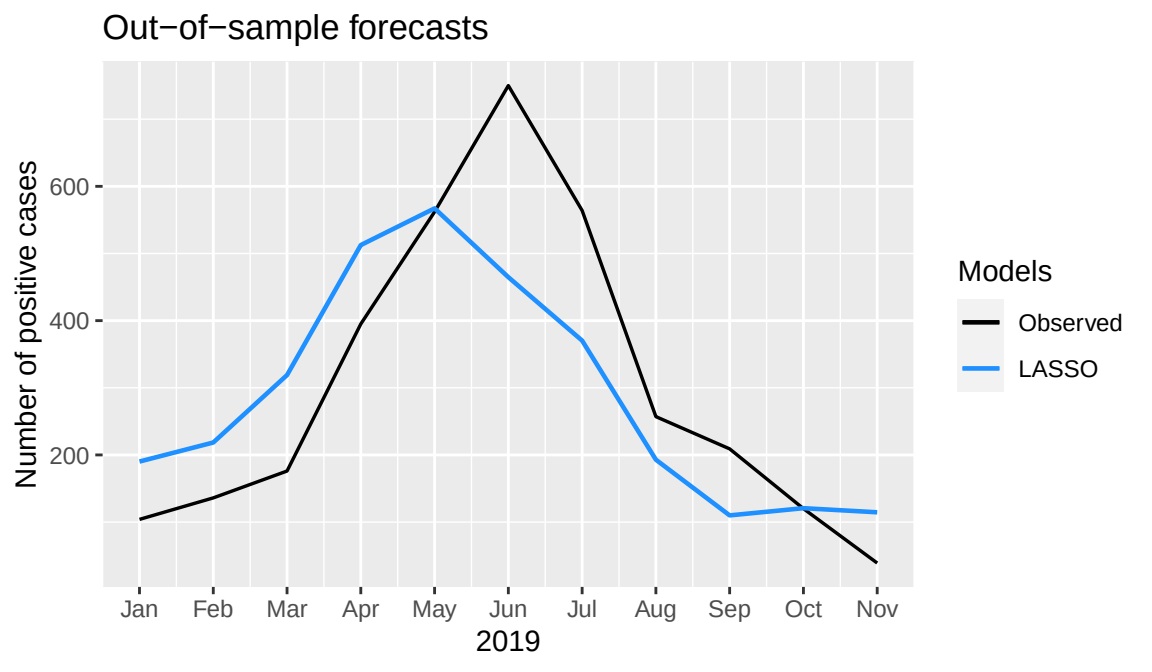

79 78 78 72 71 64 51 38 23 15 7 3 3 1
Mean−Squared Error
1e+05
4e+05
−2
0
2
4
6
Log(λ)

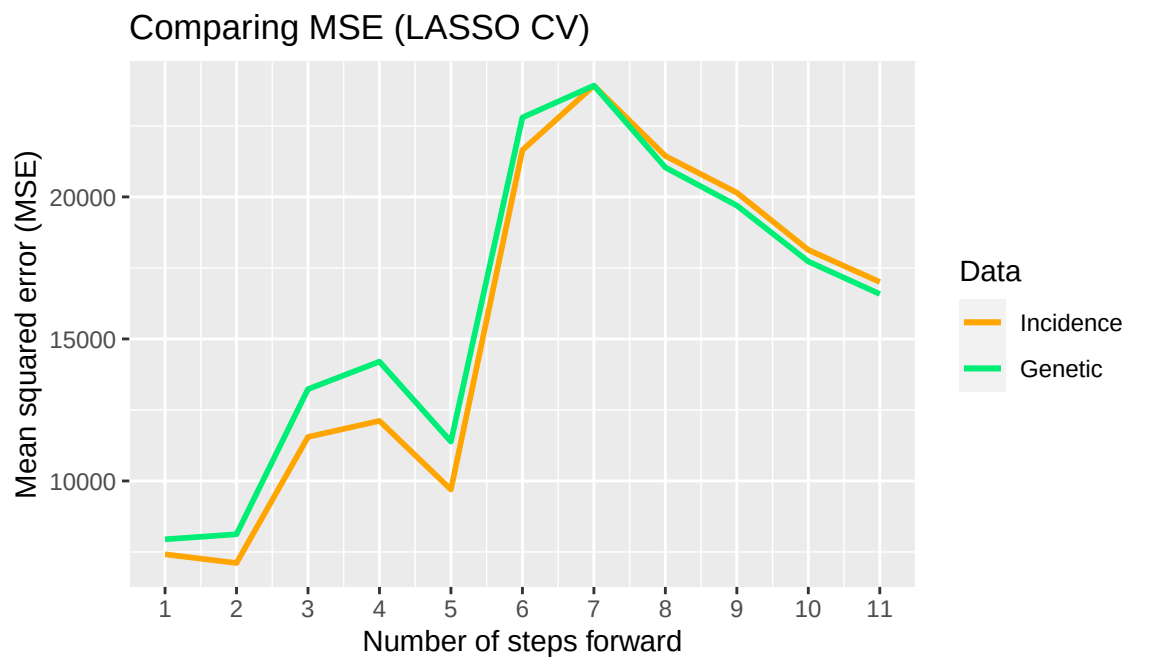

Number of positive cases
0
500
1500
2010
2012
2014
2016
2018
2020
Years

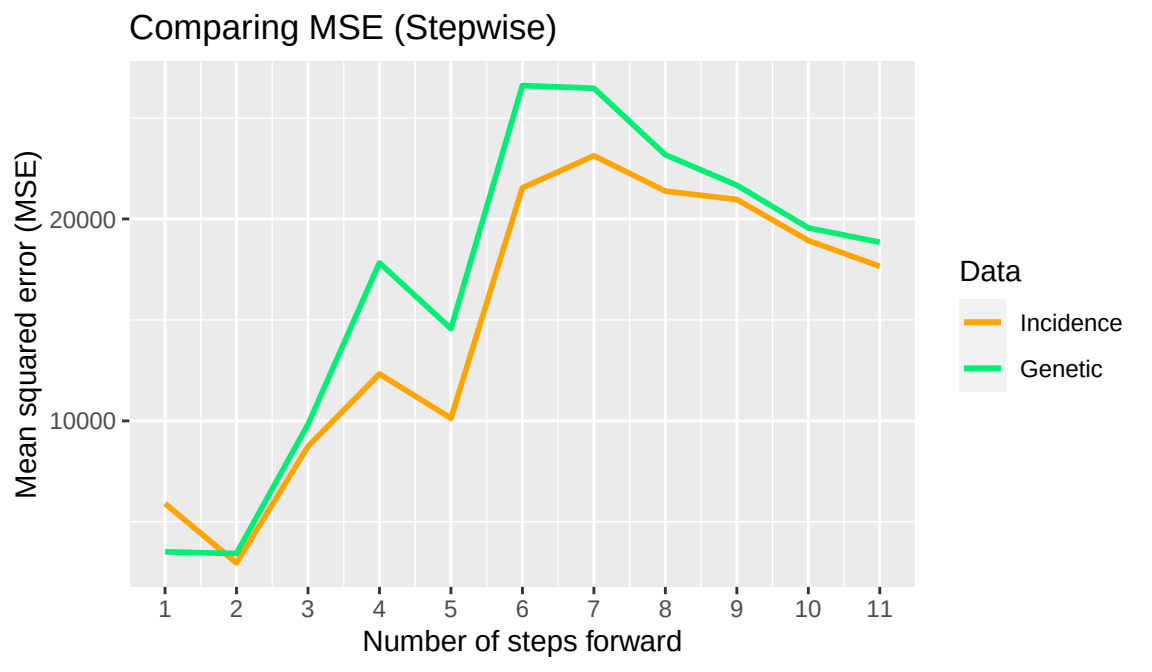

Out−of−sample forecasts
Number of positive cases
600
400
200
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
2019
Models
Observed
LASSO

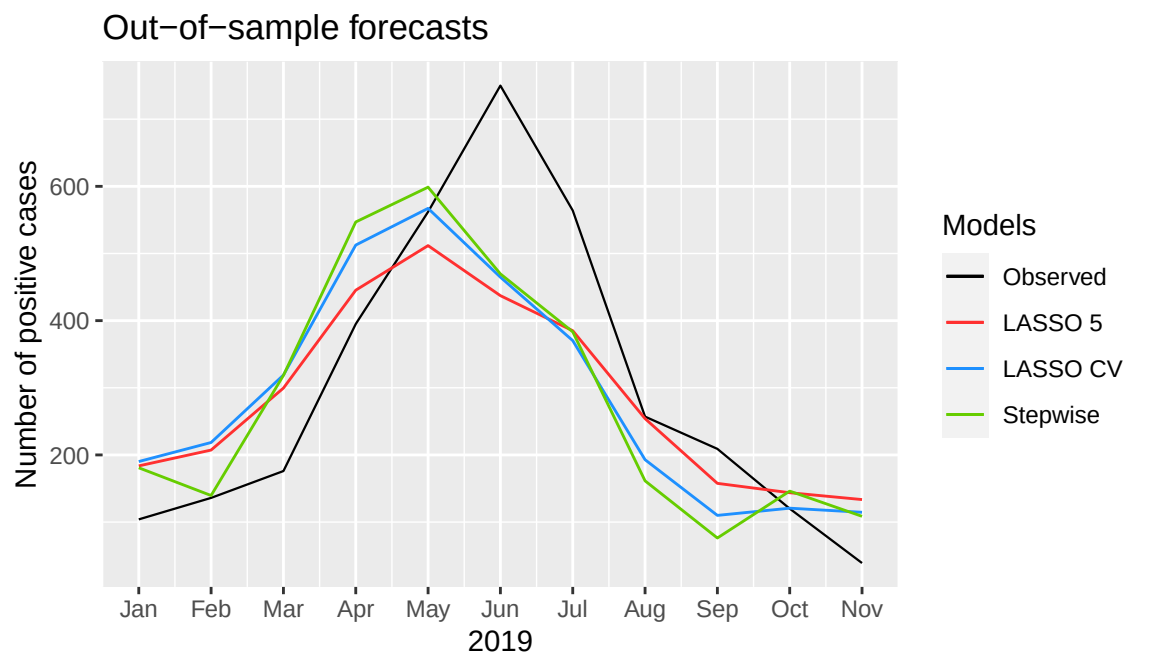

77
64
27
5
0
Coefficients
1.0
0.5
0.0
−0.5
−2
0
2
4
6
Log Lambda

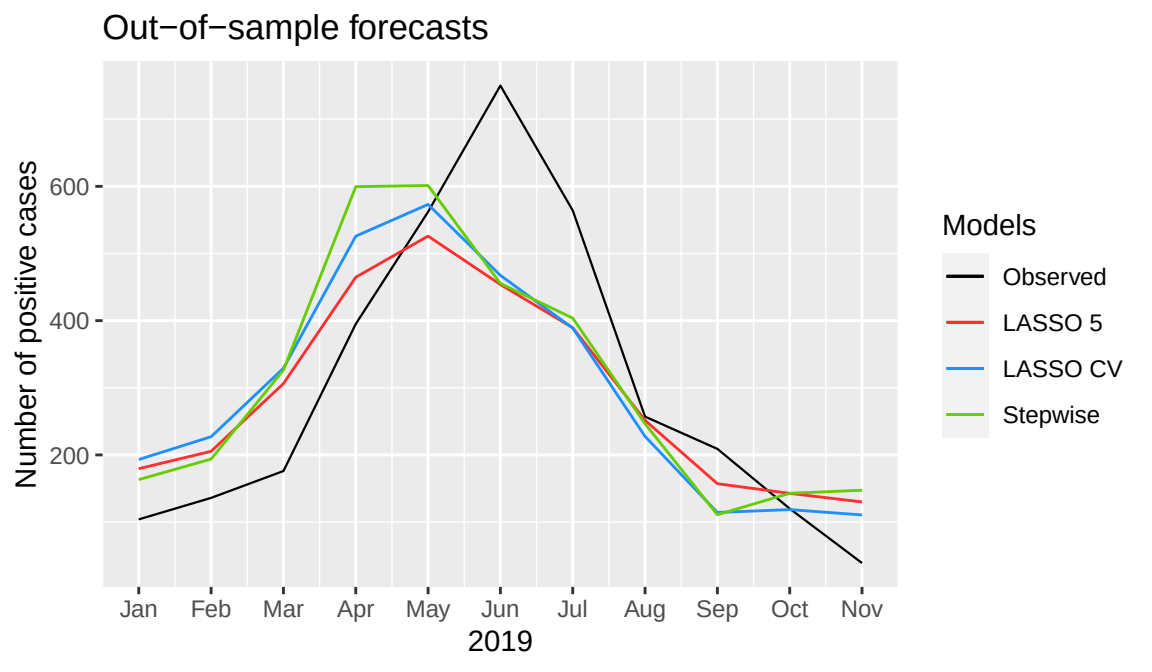

78 78 77 76 70 65 49 32 21 12 6 3 3 1
Mean−Squared Error
1e+05
3e+05
−2
0
2
4
6
Log(λ)

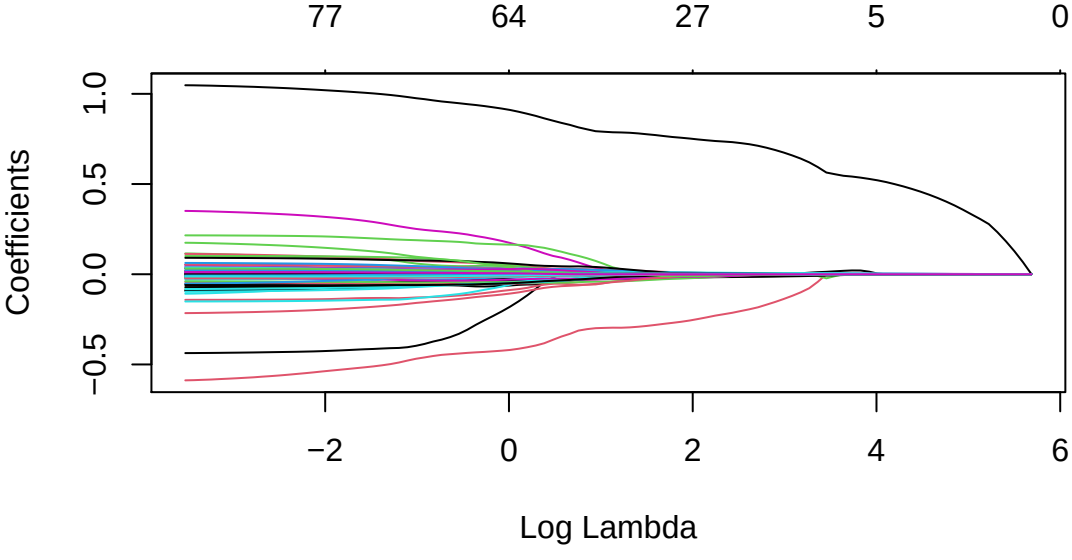

Out−of−sample forecasts
Number of positive cases
600
400
200
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
2019
Models
Observed
LASSO

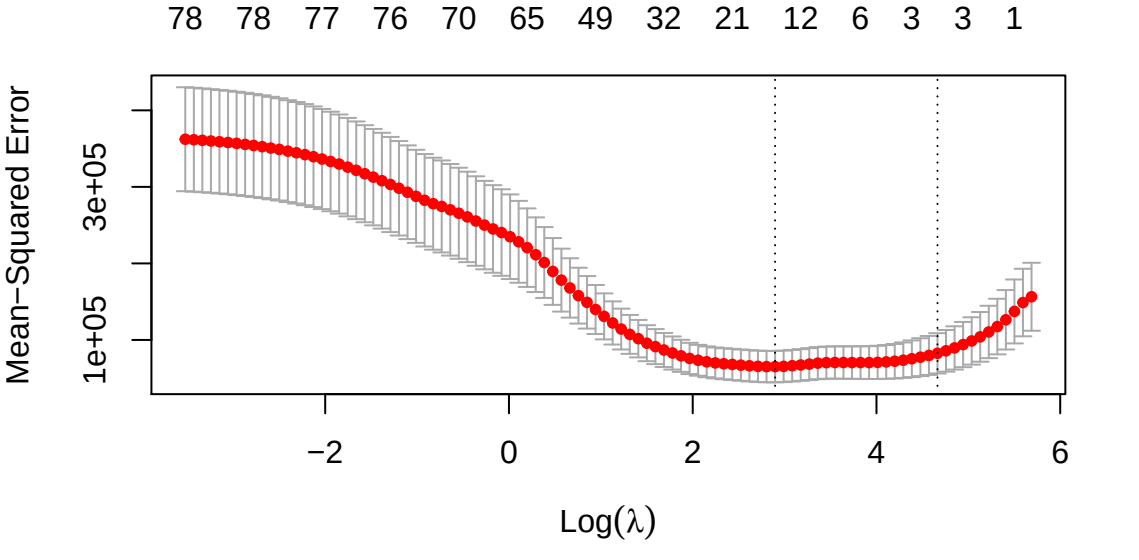

Comparing MSE (LASSO CV)
Mean squared error (MSE)
20000
15000
10000
1
2
3
4
5
6
7
8
9
10
11
Number of steps forward
Data
Incidence
Genetic

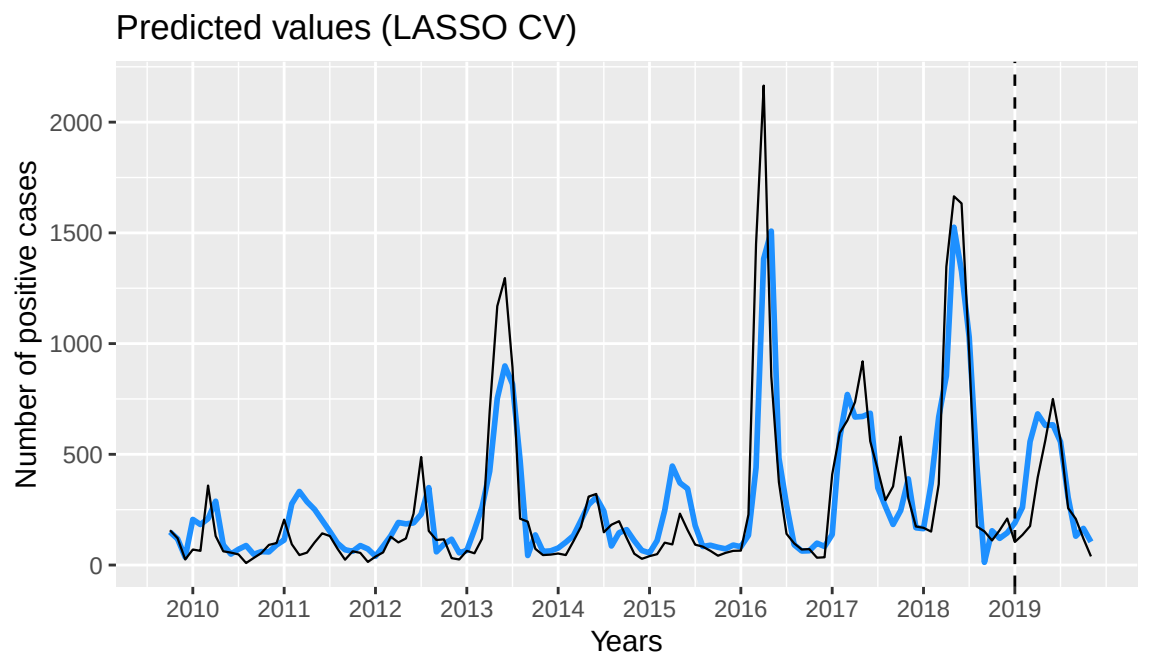

Comparing MSE (Stepwise)
Mean squared error (MSE)
20000
10000
1
2
3
4
5
6
7
8
9
10
11
Number of steps forward
Data
Incidence
Genetic

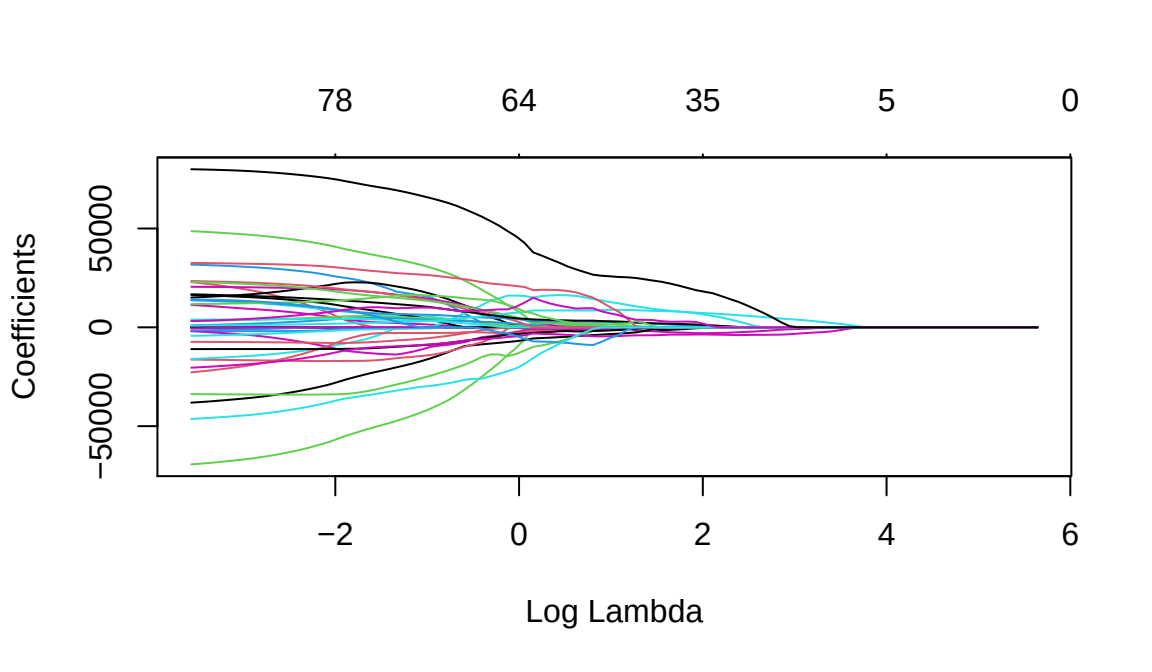

Out−of−sample forecasts
Number of positive cases
600
400
200
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
2019
Models
Observed
LASSO 5
LASSO CV
Stepwise

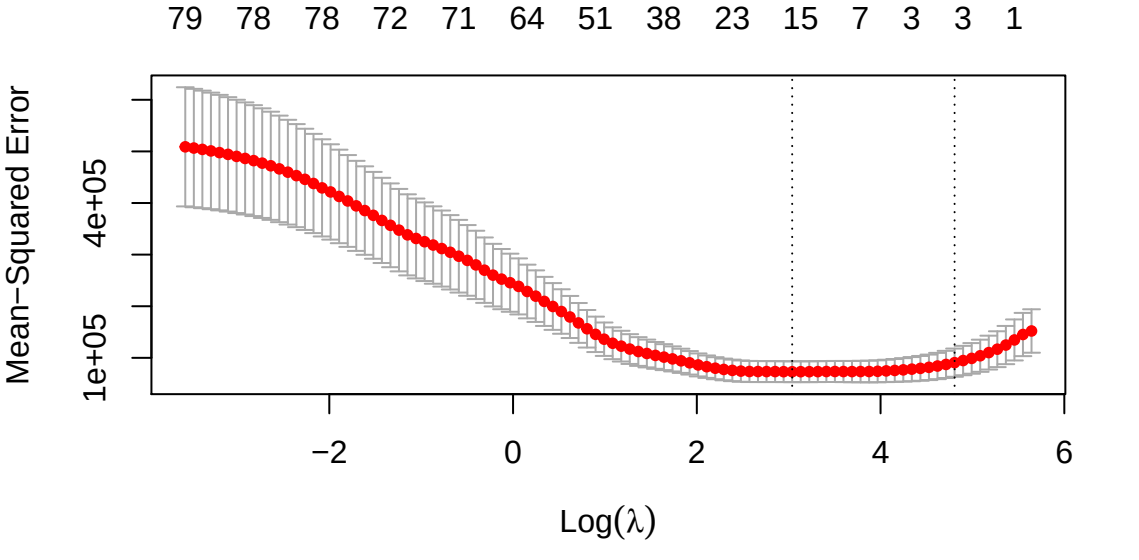

Out−of−sample forecasts
Number of positive cases
600
400
200
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
2019
Models
Observed
LASSO 5
LASSO CV
Stepwise

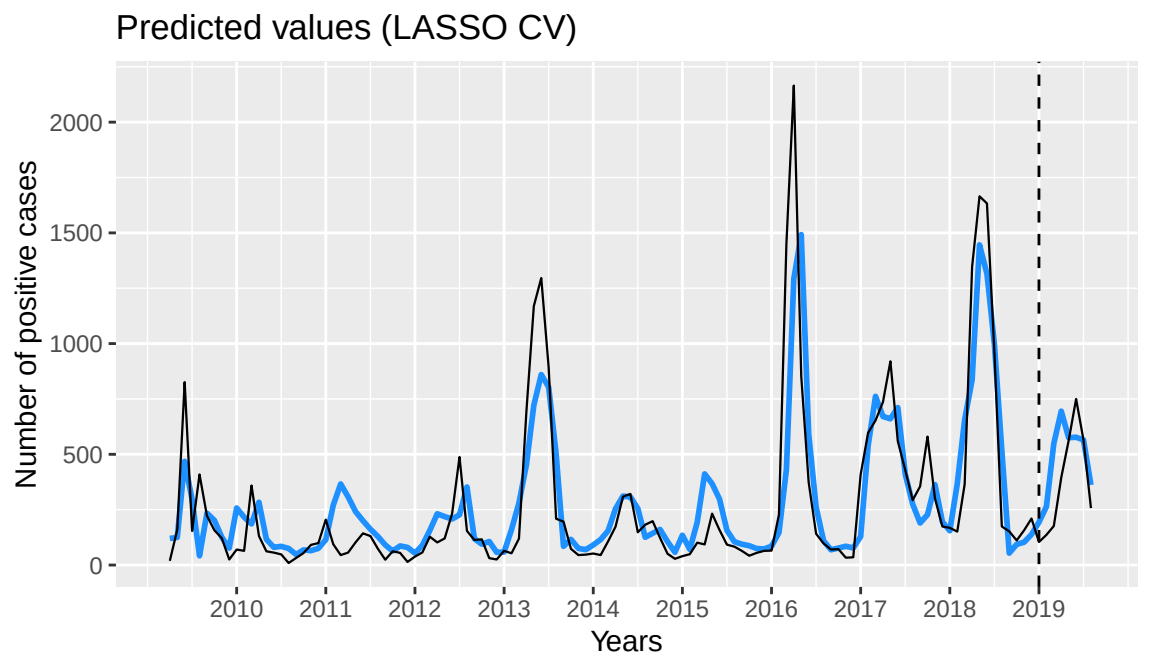

77
64
27
5
0
Coefficients
1.0
0.5
0.0
−0.5
−2
0
2
4
6
Log Lambda

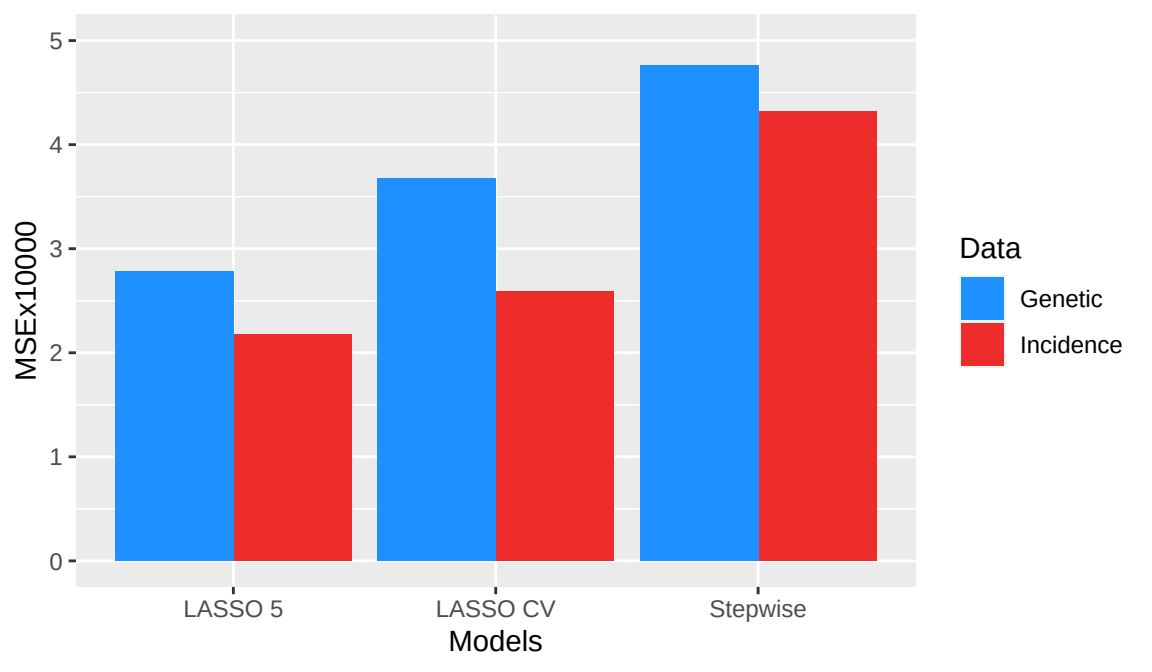

78 78 77 76 70 65 49 32 21 12 6 3 3 1
Mean−Squared Error
1e+05
3e+05
−2
0
2
4
6
Log(λ)

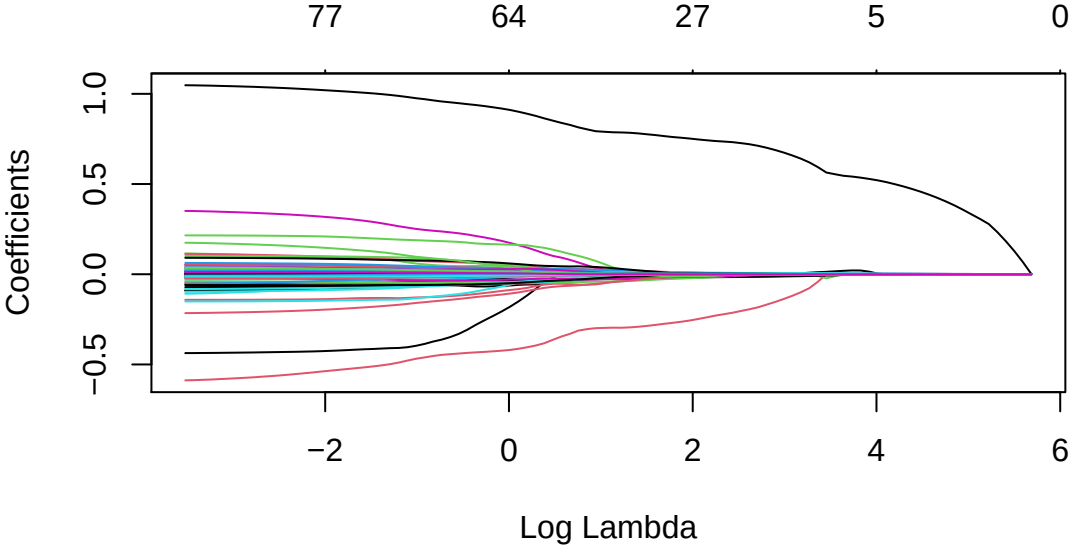

Predicted values (LASSO CV)
Number of positive cases
2000
1500
1000
500
0
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
Years

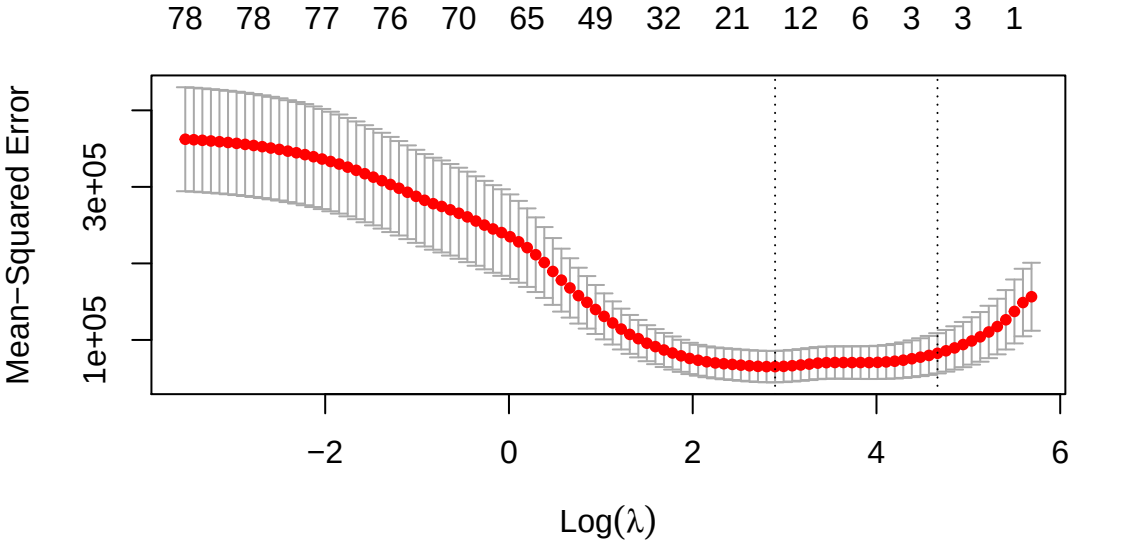

78
64
35
5
0
Coefficients
50000
0
−50000
−2
0
2
4
6
Log Lambda

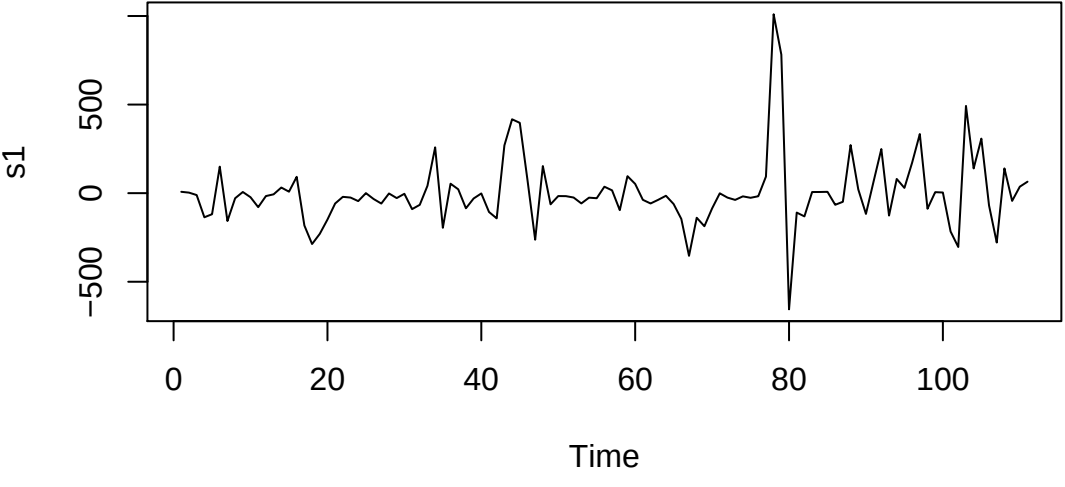

79 78 78 72 71 64 51 38 23 15 7 3 3 1
Mean−Squared Error
1e+05
4e+05
−2
0
2
4
6
Log(λ)

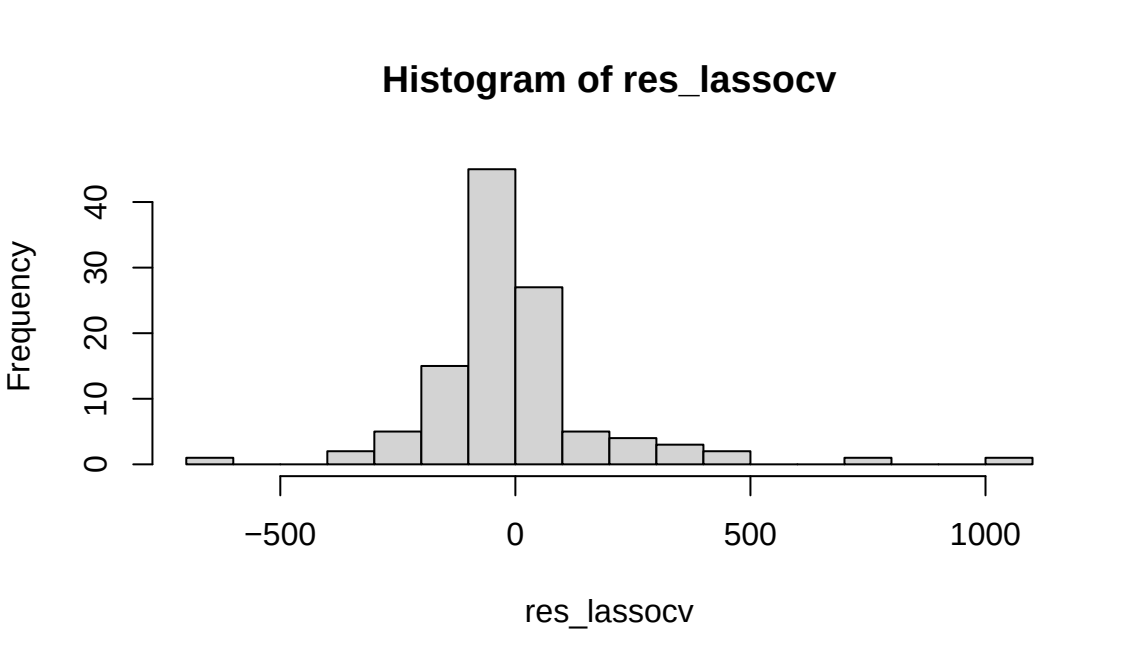

Predicted values (LASSO CV)
Number of positive cases
2000
1500
1000
500
0
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
Years

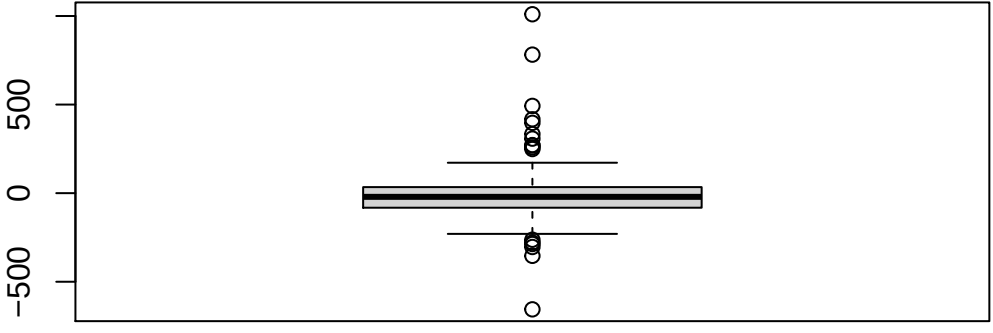

5
4
3
2
1
0
MSEx10000
LASSO 5
LASSO CV
Stepwise
Models
Data
Genetic
Incidence

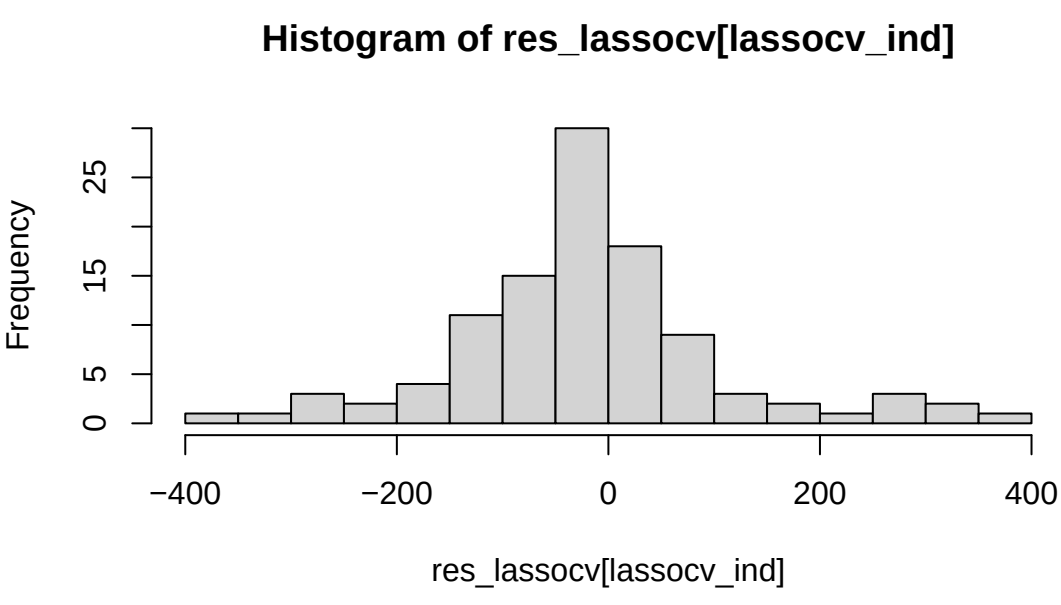

77
64
27
5
0
Coefficients
1.0
0.5
0.0
−0.5
−2
0
2
4
6
Log Lambda

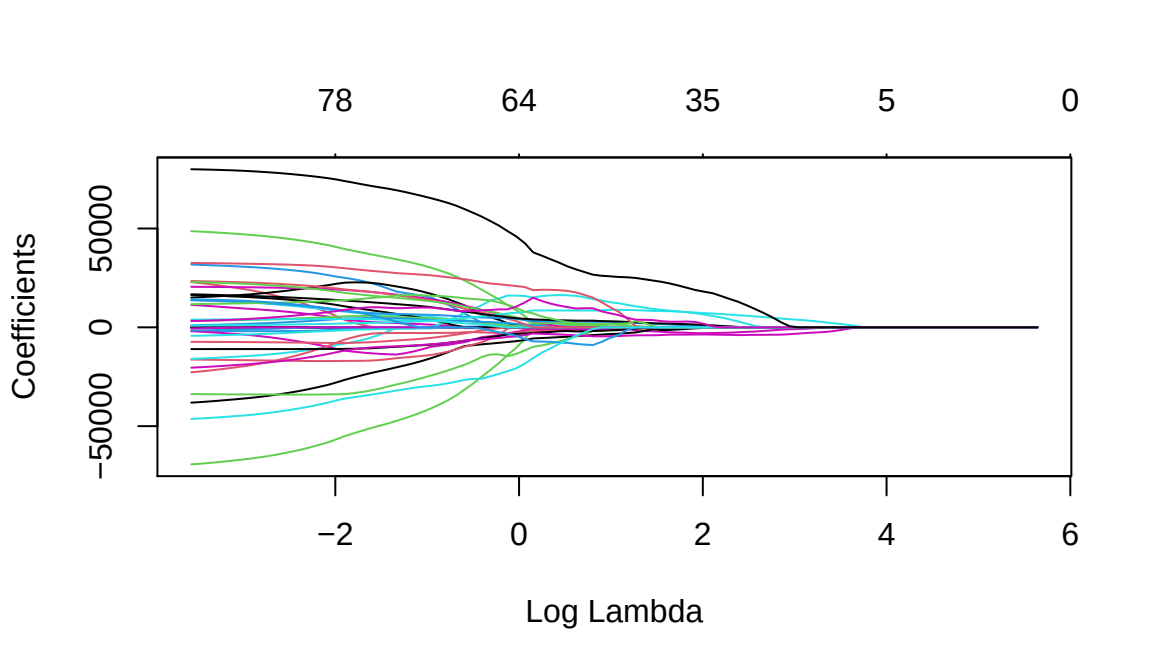

78 78 77 76 70 65 49 32 21 12 6 3 3 1
Mean−Squared Error
1e+05
3e+05
−2
0
2
4
6
Log(λ)

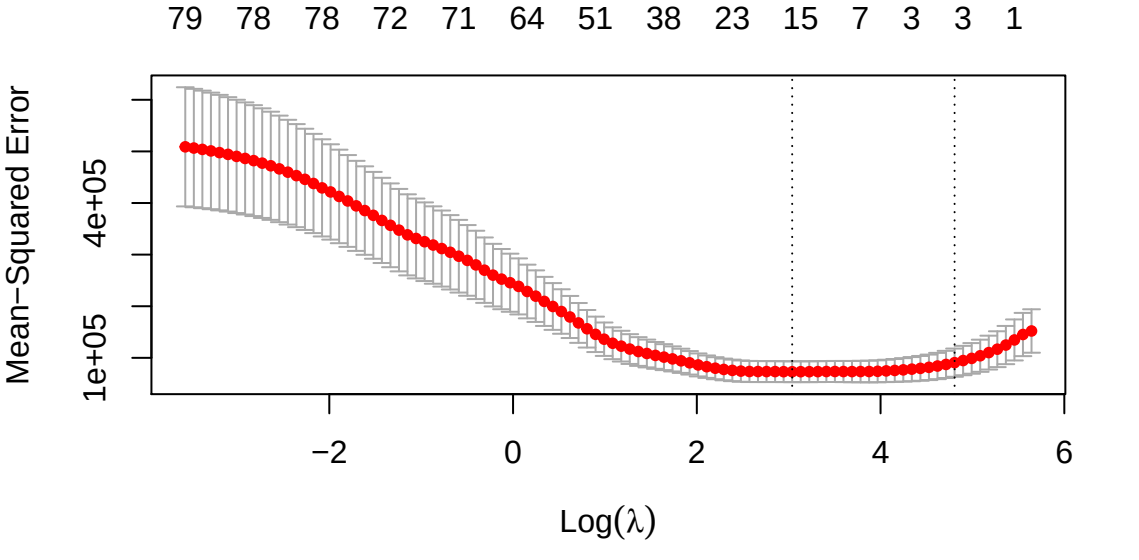

s1
500
0
−500
0
20
40
60
80
100
Time

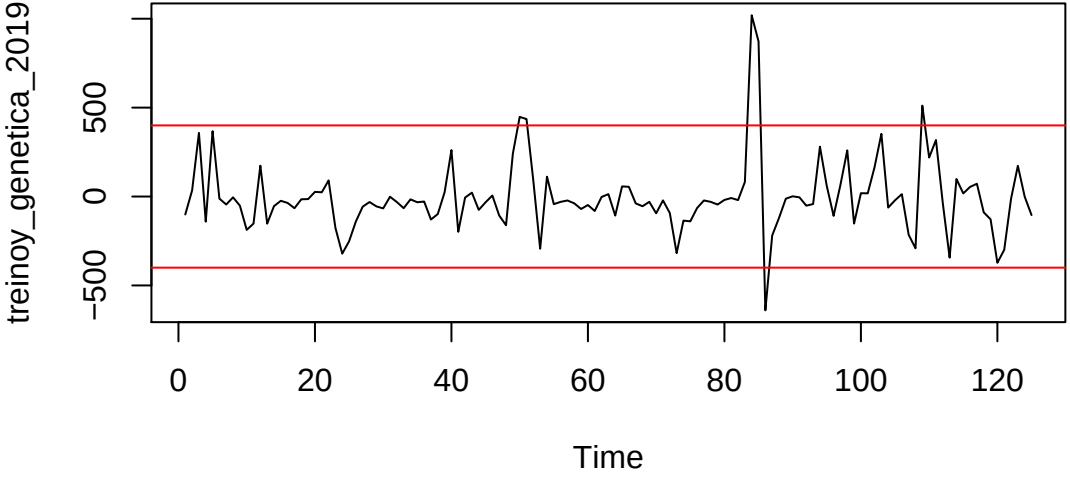

Histogram of res_lassocv
Frequency
0
10
20
30
40
−500
0
500
1000
res_lassocv

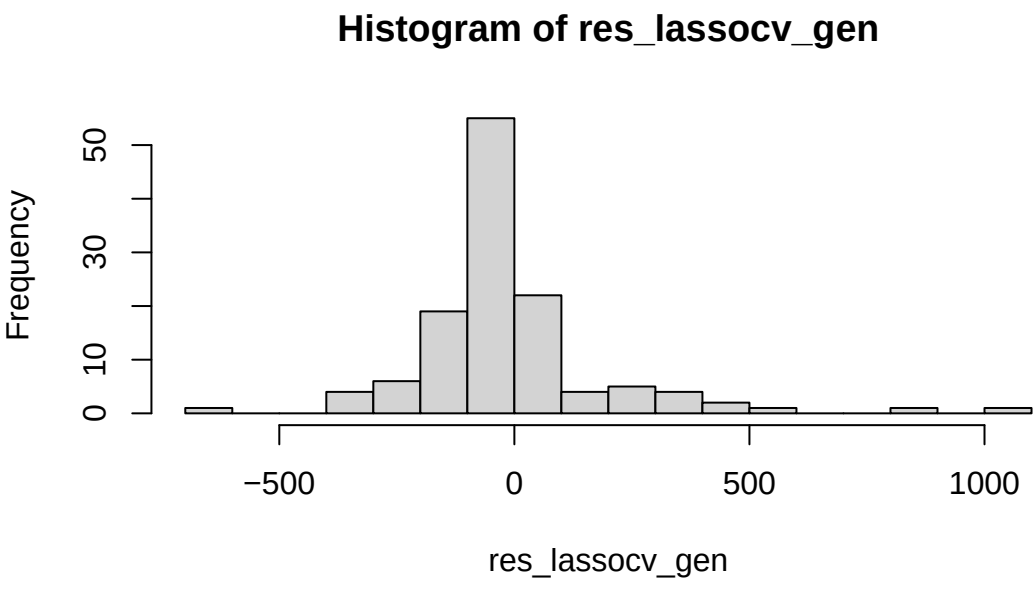

500
0
−500

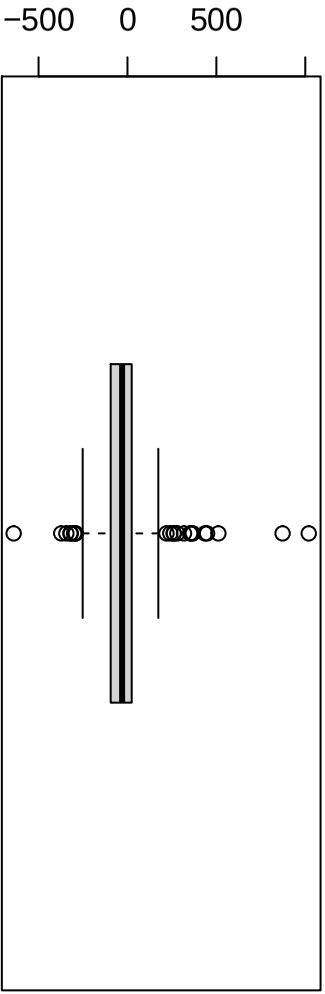

Histogram of res_lassocv[lassocv_ind]
Frequency
0
5
15
25
−400
−200
0
200
400
res_lassocv[lassocv_ind]

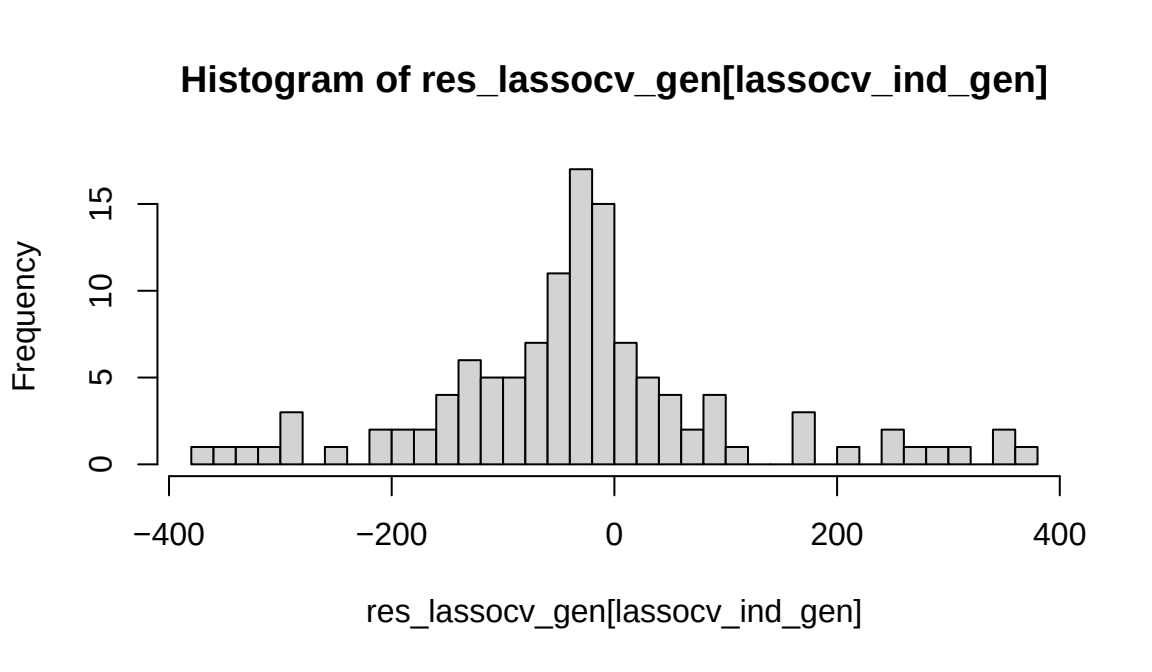

78
64
35
5
0
Coefficients
50000
0
−50000
−2
0
2
4
6
Log Lambda

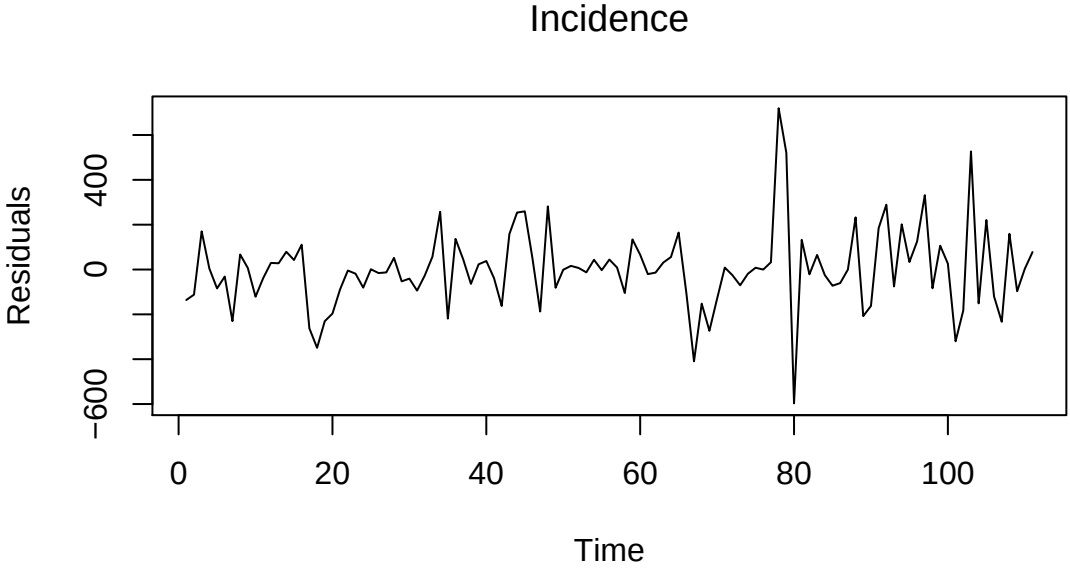

79 78 78 72 71 64 51 38 23 15 7 3 3 1
Mean−Squared Error
1e+05
4e+05
−2
0
2
4
6
Log(λ)

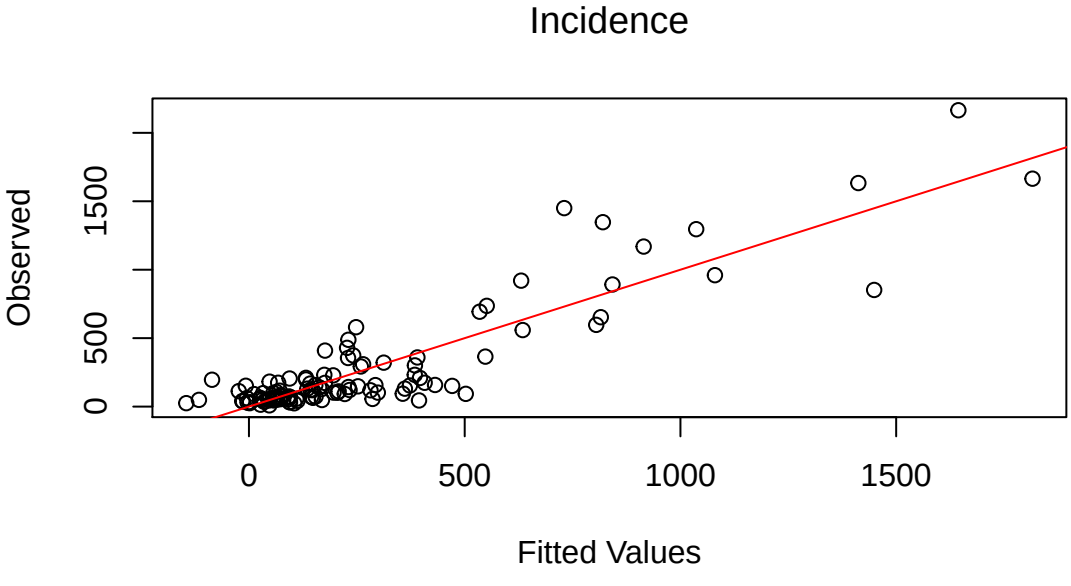

treinoy_genetica_2019
−500
0
500
0
20
40
60
80
100
120
Time

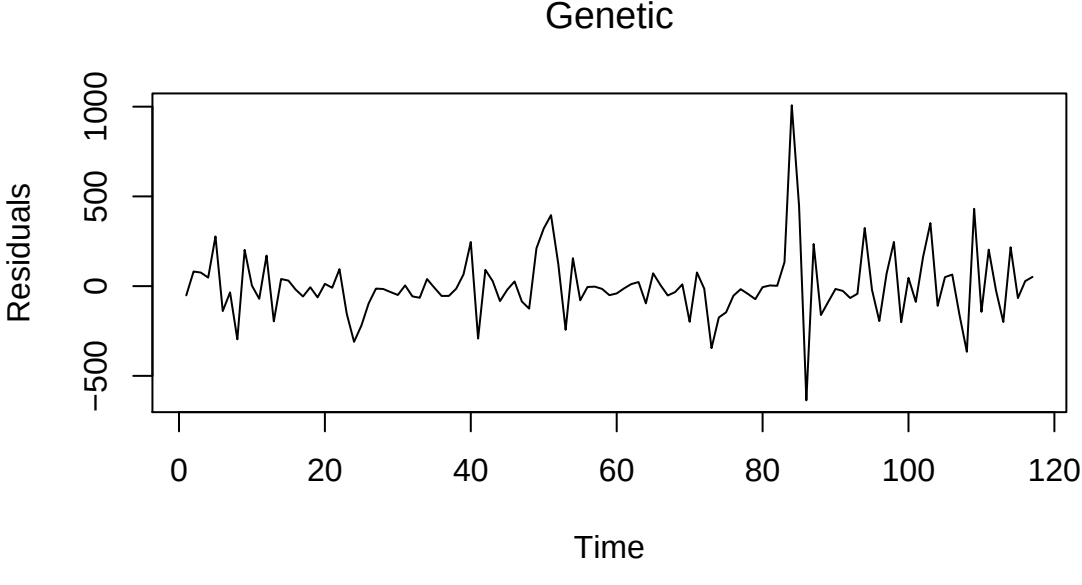

Histogram of res_lassocv_gen
Frequency
0
10
30
50
−500
0
500
1000
res_lassocv_gen

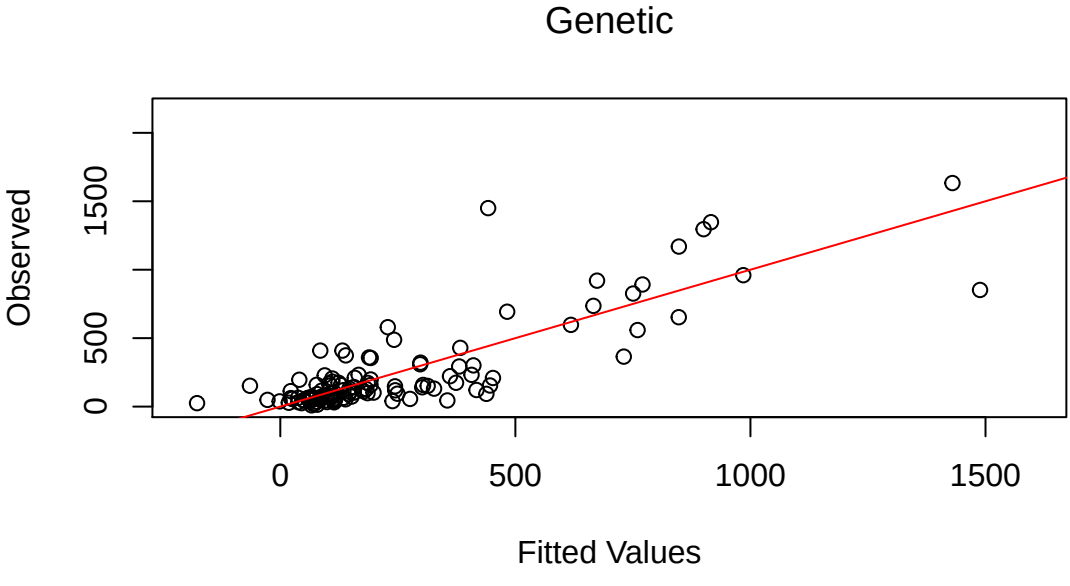

−500
0
500

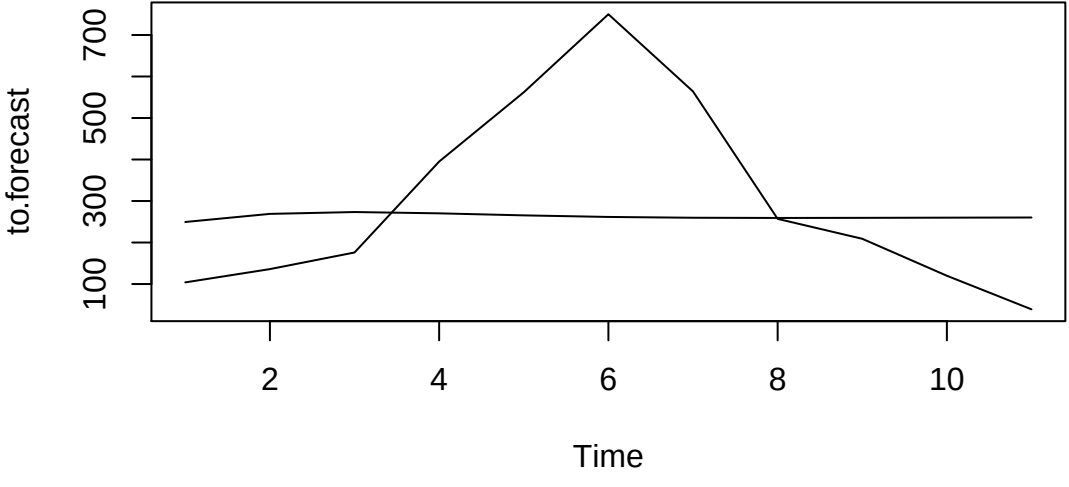

Histogram of res_lassocv_gen[lassocv_ind_gen]
Frequency
0
5
10
15
−400
−200
0
200
400
res_lassocv_gen[lassocv_ind_gen]

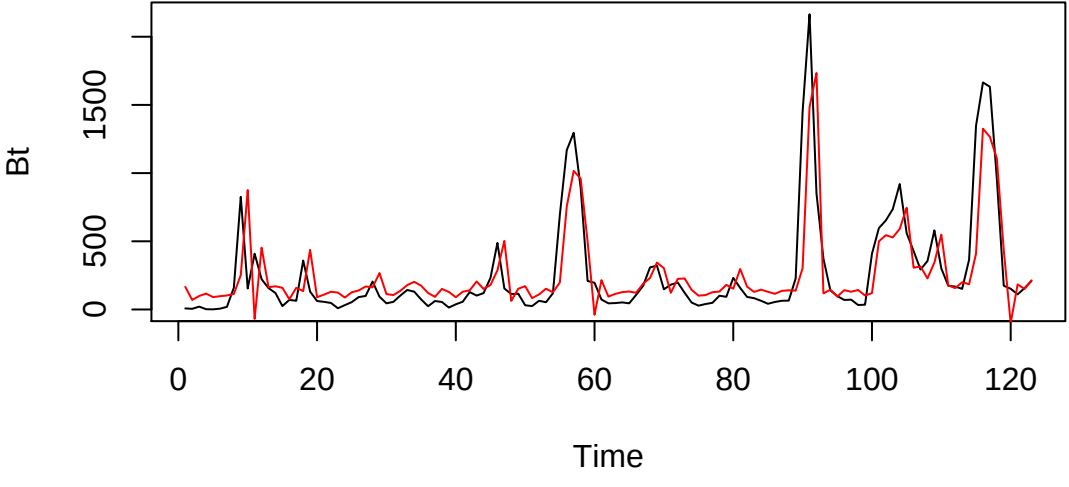

Incidence
Residuals
400
0
−600
0
20
40
60
80
100
Time

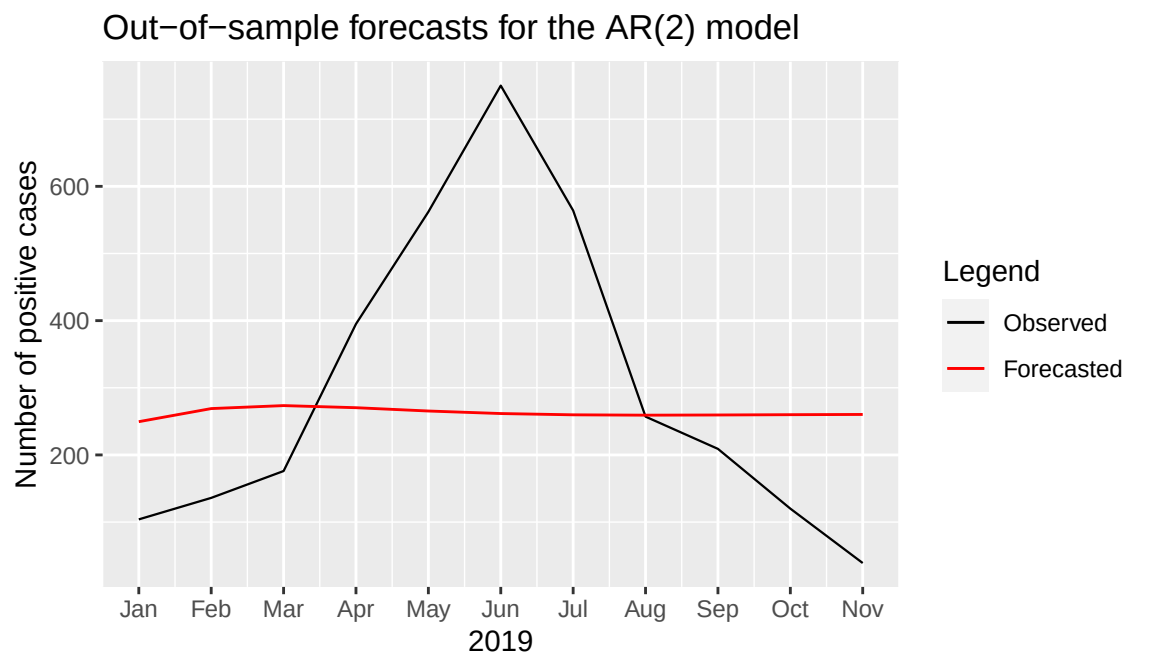

Incidence
Observed
0
500
1500
0
500
1000
1500
Fitted Values

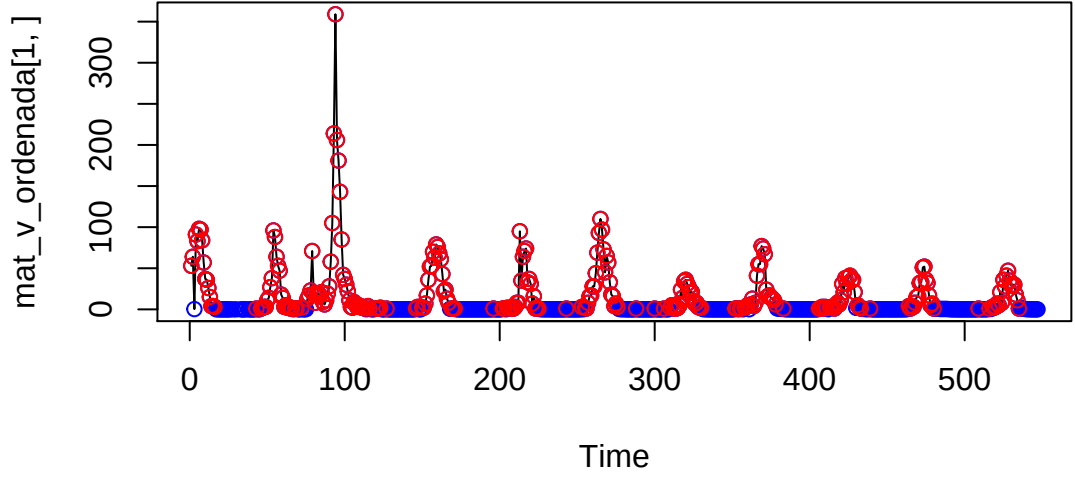

Genetic
Residuals
1000
500
0
−500
0
20
40
60
80
100
120
Time


Genetic
Observed
0
500
1500
0
500
1000
1500
Fitted Values


to.forecast
100
300
500
700
2
4
6
8
10
Time


Bt
0
500
1500
0
20
40
60
80
100
120
Time


Out−of−sample forecasts for the AR(2) model
Number of positive cases
600
400
200
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
2019
Legend
Observed
Forecasted


mat_v_ordenada[1,]
0
100
200
300
0
100
200
300
400
500
Time