Pertussis RF

Sulyok Walker

6/9/2021

#######import data################  
  
library(readxl)  
weeklyseveritygermany <- read\_excel("Downloads/weeklyseveritygermany(1).xlsx")

## New names:  
## \* `` -> ...5

data <- weeklyseveritygermany[0:74,]  
  
  
  
# load the packages  
library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.1 ──

## ✓ ggplot2 3.3.3 ✓ purrr 0.3.4  
## ✓ tibble 3.1.2 ✓ dplyr 1.0.6  
## ✓ tidyr 1.1.3 ✓ stringr 1.4.0  
## ✓ readr 1.4.0 ✓ forcats 0.5.1

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

suppressPackageStartupMessages(require(tsibble))  
suppressPackageStartupMessages(require(randomForest))  
suppressPackageStartupMessages(require(forecast))  
  
Data<-data[c(3,12, 14, 16, 18, 20, 22, 24, 26, 47)]  
dataa<-drop\_na(Data)   
  
  
  
  
# estimate the required order of differencing  
  
  
dataa<-as.ts(dataa, start=c(2020, 1), frequency=52)  
ndiffs(dataa)

## [1] 1

# log transform and difference the data  
#dataa <- log(dataa) ### makes to much NaNs  
  
dataa %>% diff(1)

## Time Series:  
## Start = 2   
## End = 74   
## Frequency = 1   
## KH school workplace publicevent gatherings publictransport stayathome  
## 2 132 0 0 0 0 0 0  
## 3 12 0 0 0 0 0 0  
## 4 4 0 0 0 0 0 0  
## 5 -15 0 0 0 0 0 0  
## 6 16 0 0 0 0 0 0  
## 7 -2 0 0 0 0 0 0  
## 8 -64 0 0 0 0 0 0  
## 9 20 2 0 1 0 0 0  
## 10 6 0 0 0 0 0 0  
## 11 -61 0 0 1 1 0 1  
## 12 0 1 2 0 3 0 1  
## 13 -49 0 0 0 0 0 0  
## 14 -5 0 0 0 0 0 0  
## 15 36 0 0 0 0 0 0  
## 16 -51 0 0 0 0 0 0  
## 17 24 0 0 0 0 0 0  
## 18 -29 0 0 0 0 0 0  
## 19 16 -1 0 0 0 0 -2  
## 20 -1 0 0 0 0 0 0  
## 21 -39 0 0 0 0 0 0  
## 22 2 0 0 0 0 0 0  
## 23 -14 0 0 0 0 0 0  
## 24 5 0 0 0 0 0 0  
## 25 9 1 0 0 0 0 0  
## 26 -17 0 0 0 0 0 0  
## 27 2 0 0 0 0 0 0  
## 28 1 -2 0 0 -1 0 0  
## 29 -7 0 0 0 0 0 0  
## 30 3 0 0 0 0 0 0  
## 31 0 0 0 0 1 0 0  
## 32 -2 1 0 0 -1 0 0  
## 33 -3 0 0 0 0 0 0  
## 34 -2 0 0 0 0 0 0  
## 35 0 0 0 0 1 0 0  
## 36 2 -1 0 0 0 0 0  
## 37 4 0 0 0 0 0 0  
## 38 -4 0 0 0 0 0 0  
## 39 15 0 0 0 0 0 0  
## 40 -14 0 0 0 0 0 0  
## 41 12 0 0 0 0 0 0  
## 42 -1 0 0 0 0 0 1  
## 43 -7 0 0 0 0 0 1  
## 44 -1 0 0 0 0 0 0  
## 45 1 0 0 0 0 0 -1  
## 46 -2 0 0 0 0 0 0  
## 47 8 0 0 0 0 0 0  
## 48 -3 0 0 0 0 0 0  
## 49 -7 0 0 0 0 1 0  
## 50 7 0 0 0 0 0 0  
## 51 -4 2 1 0 0 0 1  
## 52 -14 0 0 0 0 0 0  
## 53 6 0 0 0 0 0 0  
## 54 3 0 0 0 0 0 0  
## 55 10 0 0 0 0 0 0  
## 56 -2 0 0 0 0 0 0  
## 57 2 0 0 0 0 0 0  
## 58 -7 0 0 0 0 0 0  
## 59 6 0 0 0 0 0 0  
## 60 -11 0 0 0 0 0 0  
## 61 5 0 0 0 0 0 0  
## 62 1 0 -1 0 0 0 0  
## 63 3 0 0 0 0 0 0  
## 64 0 0 0 0 0 0 0  
## 65 -6 0 0 0 0 0 0  
## 66 -4 0 0 0 0 0 0  
## 67 -1 0 0 0 0 0 0  
## 68 7 0 0 0 0 0 0  
## 69 -4 0 0 0 0 0 0  
## 70 5 0 0 0 0 0 0  
## 71 -3 0 0 0 0 0 0  
## 72 -7 0 0 0 0 0 0  
## 73 7 0 0 0 0 0 0  
## 74 3 0 0 0 0 0 0  
## internalmovement internationaltravel StringencyIndex  
## 2 0 0 0.00  
## 3 0 0 0.00  
## 4 0 0 5.56  
## 5 0 0 0.00  
## 6 0 0 0.00  
## 7 0 0 5.55  
## 8 0 0 0.00  
## 9 0 1 13.89  
## 10 0 0 0.00  
## 11 0 0 7.87  
## 12 2 3 43.98  
## 13 0 0 0.00  
## 14 0 0 0.00  
## 15 0 0 0.00  
## 16 0 0 0.00  
## 17 0 0 0.00  
## 18 0 0 -1.39  
## 19 0 0 -11.11  
## 20 0 -1 -2.78  
## 21 0 0 -1.85  
## 22 0 0 0.00  
## 23 0 0 0.00  
## 24 0 0 0.00  
## 25 0 0 3.71  
## 26 0 0 0.00  
## 27 0 0 0.00  
## 28 0 0 -8.34  
## 29 0 0 0.00  
## 30 0 0 0.00  
## 31 0 0 2.78  
## 32 0 0 -0.93  
## 33 0 0 0.00  
## 34 0 0 0.00  
## 35 0 0 2.78  
## 36 -2 0 -10.18  
## 37 0 0 0.00  
## 38 0 0 0.00  
## 39 0 0 0.00  
## 40 0 0 0.00  
## 41 0 0 0.00  
## 42 1 0 9.26  
## 43 0 0 1.85  
## 44 0 0 0.00  
## 45 0 0 1.39  
## 46 1 0 2.77  
## 47 0 0 0.00  
## 48 0 0 0.00  
## 49 -1 0 2.78  
## 50 0 0 0.00  
## 51 0 0 14.82  
## 52 0 0 0.00  
## 53 0 0 0.00  
## 54 1 0 0.92  
## 55 0 0 0.00  
## 56 0 0 0.00  
## 57 0 0 0.00  
## 58 0 0 0.00  
## 59 0 0 0.00  
## 60 0 0 0.00  
## 61 0 0 -1.85  
## 62 0 0 -3.70  
## 63 0 0 0.00  
## 64 -1 0 -2.78  
## 65 0 0 0.00  
## 66 0 0 0.00  
## 67 0 0 0.00  
## 68 0 0 0.00  
## 69 0 0 0.00  
## 70 0 0 0.00  
## 71 0 0 0.00  
## 72 0 0 0.00  
## 73 0 0 0.00  
## 74 0 0 0.00

lag\_order <- 6 # the desired number of lags (six weeks)  
horizon <- 12 # the forecast horizon (twelve weeks)  
  
d\_mbd <- embed(dataa, lag\_order + 1)  
dframe<-as.data.frame(d\_mbd)  
  
##########train test split#############################  
  
test<-dframe[54:68,]   
dframe<-dframe[1:53,] #from 2020 feb 9 to 2021 feb 7  
  
  
 # fit the model  
library(randomForestSRC)

##   
## randomForestSRC 2.11.0   
##   
## Type rfsrc.news() to see new features, changes, and bug fixes.   
##

##   
## Attaching package: 'randomForestSRC'

## The following object is masked from 'package:purrr':  
##   
## partial

set.seed(123)  
tunedgrow<-tune.rfsrc(V1 ~. , data=dframe, ntree=50, nodesize=2, mtry=3, trace=TRUE, na.action="na.omit")

## nodesize = 2 mtry = 3 OOB error = 8.35%   
## Searching left ...  
## nodesize = 2 mtry = 2 OOB error = 9.36%   
## -0.1207318 0.001   
## nodesize = 2 mtry = 1 OOB error = 10.76%   
## -0.2878399 0.001   
## Searching right ...  
## nodesize = 2 mtry = 4 OOB error = 9.65%   
## -0.1558423 0.001   
## nodesize = 2 mtry = 5 OOB error = 8.79%   
## -0.05227581 0.001   
## nodesize = 2 mtry = 6 OOB error = 8.7%   
## -0.04159229 0.001   
## nodesize = 2 mtry = 7 OOB error = 9.95%   
## -0.1918105 0.001

tunedgrow

## $results  
## nodesize mtry err  
## 1 2 1 0.10755839  
## 2 2 2 0.09360178  
## 3 2 3 0.08351845  
## 4 2 4 0.09653415  
## 5 2 5 0.08788444  
## 6 2 6 0.08699217  
## 7 2 7 0.09953816  
##   
## $optimal  
## nodesize mtry   
## 2 3   
##   
## $rf  
## Sample size: 53  
## Number of trees: 500  
## Forest terminal node size: 2  
## Average no. of terminal nodes: 14.436  
## No. of variables tried at each split: 3  
## Total no. of variables: 69  
## Resampling used to grow trees: swor  
## Resample size used to grow trees: 33  
## Analysis: RF-R  
## Family: regr  
## Splitting rule: mse \*random\*  
## Number of random split points: 10  
## % variance explained: 92.05  
## Error rate: 223.06

set.seed(123)   
grow<-rfsrc(V1 ~. , data=dframe, ntree=50, nodesize=2, mtry=3, trace=TRUE, na.action="na.omit")  
grow

## Sample size: 53  
## Number of trees: 50  
## Forest terminal node size: 2  
## Average no. of terminal nodes: 14.7  
## No. of variables tried at each split: 3  
## Total no. of variables: 69  
## Resampling used to grow trees: swor  
## Resample size used to grow trees: 33  
## Analysis: RF-R  
## Family: regr  
## Splitting rule: mse \*random\*  
## Number of random split points: 10  
## % variance explained: 91.65  
## Error rate: 234.28

library(ggRandomForests)

##   
## Attaching package: 'ggRandomForests'

## The following object is masked from 'package:randomForestSRC':  
##   
## partial.rfsrc

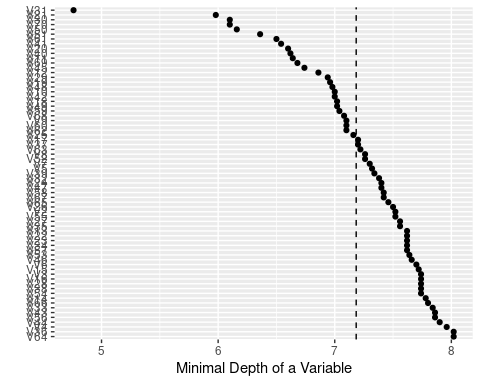
gg\_md <- gg\_minimal\_depth(grow)

## minimal depth variable selection ...  
##   
##   
## -----------------------------------------------------------  
## family : regr   
## var. selection : Minimal Depth   
## conservativeness : medium   
## x-weighting used? : TRUE   
## dimension : 69   
## sample size : 53   
## ntree : 50   
## nsplit : 10   
## mtry : 3   
## nodesize : 2   
## refitted forest : FALSE   
## model size : 27   
## depth threshold : 7.1839   
## PE (true OOB) : 234.28   
##   
##   
## Top variables:  
## depth vimp  
## V31 4.76 NA  
## V21 5.98 NA  
## V30 6.10 NA  
## V70 6.10 NA  
## V50 6.16 NA  
## V51 6.36 NA  
## V61 6.50 NA  
## V41 6.54 NA  
## V20 6.60 NA  
## V40 6.62 NA  
## V11 6.64 NA  
## V69 6.68 NA  
## V45 6.74 NA  
## V12 6.86 NA  
## V29 6.94 NA  
## V19 6.96 NA  
## V48 6.98 NA  
## V10 7.00 NA  
## V42 7.00 NA  
## V18 7.02 NA  
## V49 7.02 NA  
## V59 7.04 NA  
## V68 7.08 NA  
## V9 7.10 NA  
## V60 7.10 NA  
## V62 7.10 NA  
## V25 7.16 NA  
## -----------------------------------------------------------

md<-plot(gg\_md)

## Coordinate system already present. Adding new coordinate system, which will replace the existing one.

md



pdf(file = "md",   
 width = 5,   
 height = 8)   
plot(md)  
dev.off()

## png   
## 2

set.seed(123)  
vimps<-gg\_vimp(grow)

## Warning in gg\_vimp.rfsrc(grow): rfsrc object does not contain VIMP information.  
## Calculating...

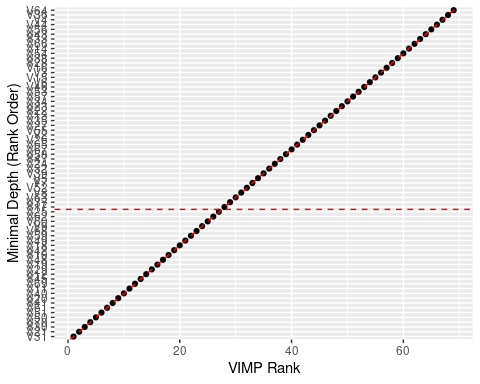
vimps

## vars set vimp positive  
## 1 V30 VIMP 1.580671e+02 TRUE  
## 2 V13 VIMP 1.092677e+02 TRUE  
## 3 V69 VIMP 1.019292e+02 TRUE  
## 4 V21 VIMP 9.143309e+01 TRUE  
## 5 V11 VIMP 8.587445e+01 TRUE  
## 6 V31 VIMP 8.397625e+01 TRUE  
## 7 V34 VIMP 8.070061e+01 TRUE  
## 8 V49 VIMP 7.716107e+01 TRUE  
## 9 V25 VIMP 6.254656e+01 TRUE  
## 10 V51 VIMP 6.098546e+01 TRUE  
## 11 V24 VIMP 5.857404e+01 TRUE  
## 12 V45 VIMP 5.407799e+01 TRUE  
## 13 V60 VIMP 4.606250e+01 TRUE  
## 14 V59 VIMP 4.505243e+01 TRUE  
## 15 V29 VIMP 3.548397e+01 TRUE  
## 16 V23 VIMP 3.530048e+01 TRUE  
## 17 V41 VIMP 3.277711e+01 TRUE  
## 18 V9 VIMP 2.791555e+01 TRUE  
## 19 V56 VIMP 2.320301e+01 TRUE  
## 20 V54 VIMP 2.267355e+01 TRUE  
## 21 V39 VIMP 2.221573e+01 TRUE  
## 22 V65 VIMP 2.095429e+01 TRUE  
## 23 V19 VIMP 2.049354e+01 TRUE  
## 24 V14 VIMP 1.999996e+01 TRUE  
## 25 V63 VIMP 1.640420e+01 TRUE  
## 26 V50 VIMP 1.613347e+01 TRUE  
## 27 V62 VIMP 1.466666e+01 TRUE  
## 28 V27 VIMP 1.337326e+01 TRUE  
## 29 V42 VIMP 1.252286e+01 TRUE  
## 30 V18 VIMP 1.091952e+01 TRUE  
## 31 V22 VIMP 9.349576e+00 TRUE  
## 32 V17 VIMP 9.315026e+00 TRUE  
## 33 V55 VIMP 9.009463e+00 TRUE  
## 34 V6 VIMP 7.515167e+00 TRUE  
## 35 V32 VIMP 5.738091e+00 TRUE  
## 36 V12 VIMP 4.596496e+00 TRUE  
## 37 V37 VIMP 4.481858e+00 TRUE  
## 38 V52 VIMP 4.234059e+00 TRUE  
## 39 V53 VIMP 3.965051e+00 TRUE  
## 40 V48 VIMP 1.738089e+00 TRUE  
## 41 V5 VIMP 1.592972e+00 TRUE  
## 42 V43 VIMP 1.537579e+00 TRUE  
## 43 V33 VIMP 1.468563e+00 TRUE  
## 44 V40 VIMP 1.236898e+00 TRUE  
## 45 V20 VIMP 1.196832e+00 TRUE  
## 46 V57 VIMP 9.229000e-01 TRUE  
## 47 V38 VIMP 7.966106e-01 TRUE  
## 48 V26 VIMP 6.484674e-01 TRUE  
## 49 V47 VIMP 4.875378e-01 TRUE  
## 50 V8 VIMP 4.124313e-01 TRUE  
## 51 V58 VIMP 4.044946e-01 TRUE  
## 52 V36 VIMP 1.845195e-01 TRUE  
## 53 V15 VIMP 8.487589e-02 TRUE  
## 54 V2 VIMP 3.153180e-02 TRUE  
## 55 V68 VIMP 2.640259e-02 TRUE  
## 56 V66 VIMP 5.463459e-03 TRUE  
## 57 V64 VIMP 2.273737e-14 TRUE  
## 58 V3 VIMP -1.006465e-01 FALSE  
## 59 V46 VIMP -1.490612e-01 FALSE  
## 60 V16 VIMP -2.478793e-01 FALSE  
## 61 V35 VIMP -3.363476e-01 FALSE  
## 62 V44 VIMP -8.783281e-01 FALSE  
## 63 V10 VIMP -1.020693e+00 FALSE  
## 64 V28 VIMP -1.084111e+00 FALSE  
## 65 V61 VIMP -1.512069e+00 FALSE  
## 66 V4 VIMP -2.927358e+00 FALSE  
## 67 V70 VIMP -3.247817e+00 FALSE  
## 68 V67 VIMP -3.552290e+00 FALSE  
## 69 V7 VIMP -6.180874e+00 FALSE

pdf(file = "vimp",   
 width = 5,   
 height = 8)   
plot(vimps)  
dev.off()

## png   
## 2

plot(gg\_minimal\_vimp(gg\_md))



set.seed(12345)  
pred<-predict(grow, test)  
pred

## Sample size of test (predict) data: 15  
## Number of grow trees: 50  
## Average no. of grow terminal nodes: 14.7  
## Total no. of grow variables: 69  
## Resampling used to grow trees: swor  
## Resample size used to grow trees: 9  
## Analysis: RF-R  
## Family: regr  
## % variance explained: -226.45  
## Test set error rate: 48.72

pred$predicted

## [1] 17.61793 18.80060 17.20317 20.42983 19.17417 18.10083 18.13083 16.90417  
## [9] 17.88083 18.27417 18.89417 18.26083 18.89750 18.26083 18.62750

RMSE = function(m, o){  
 sqrt(mean((m - o)^2))  
}  
  
RMSE(pred$predicted, test$V1)

## [1] 6.979877

###### PD plots ################################  
  
par(mar=c(1, 1, 1, 1))  
  
  
pdf(file = "partial",   
 width = 15,   
 height = 15)   
plot.variable(grow, partial = TRUE, smooth.lines = TRUE)

## Warning in bxp(z = list(stats = structure(c(40.2983946278068, 40.943541444276, :  
## some notches went outside hinges ('box'): maybe set notch=FALSE

## Warning in bxp(z = list(stats = structure(c(41.934354850307, 43.0248658420109, :  
## some notches went outside hinges ('box'): maybe set notch=FALSE

## Warning in bxp(z = list(stats = structure(c(40.1491590953521,  
## 40.8858796211203, : some notches went outside hinges ('box'): maybe set  
## notch=FALSE

## Warning in bxp(z = list(stats = structure(c(42.5061609512907, 43.64332542163, :  
## some notches went outside hinges ('box'): maybe set notch=FALSE

## Warning in bxp(z = list(stats = structure(c(40.2621520056232,  
## 40.9988725313914, : some notches went outside hinges ('box'): maybe set  
## notch=FALSE

## Warning in bxp(z = list(stats = structure(c(40.2589047701812,  
## 40.9545992335731, : some notches went outside hinges ('box'): maybe set  
## notch=FALSE

## Warning in bxp(z = list(stats = structure(c(40.1085247283325,  
## 40.8287789582859, : some notches went outside hinges ('box'): maybe set  
## notch=FALSE

## Warning in bxp(z = list(stats = structure(c(43.0170441350599,  
## 43.5831306792853, : some notches went outside hinges ('box'): maybe set  
## notch=FALSE

## Warning in bxp(z = list(stats = structure(c(40.215440088674, 40.9521606144422, :  
## some notches went outside hinges ('box'): maybe set notch=FALSE

## Warning in bxp(z = list(stats = structure(c(42.9833100315329,  
## 43.5045356426211, : some notches went outside hinges ('box'): maybe set  
## notch=FALSE

## Warning in bxp(z = list(stats = structure(c(45.1915987939724,  
## 46.4093219966416, : some notches went outside hinges ('box'): maybe set  
## notch=FALSE

## Warning in bxp(z = list(stats = structure(c(40.0894438046229,  
## 40.8252696945633, : some notches went outside hinges ('box'): maybe set  
## notch=FALSE

## Warning in bxp(z = list(stats = structure(c(40.1854300423282,  
## 40.8878104271093, : some notches went outside hinges ('box'): maybe set  
## notch=FALSE

## Warning in bxp(z = list(stats = structure(c(39.9351346074156,  
## 40.6743691034619, : some notches went outside hinges ('box'): maybe set  
## notch=FALSE

## Warning in bxp(z = list(stats = structure(c(39.8227937625084,  
## 40.5350902965871, : some notches went outside hinges ('box'): maybe set  
## notch=FALSE

## Warning in bxp(z = list(stats = structure(c(45.0090850815861, 46.593555291418, :  
## some notches went outside hinges ('box'): maybe set notch=FALSE

## Warning in bxp(z = list(stats = structure(c(40.3007052970683, 41.038438972619, :  
## some notches went outside hinges ('box'): maybe set notch=FALSE

## Warning in bxp(z = list(stats = structure(c(41.6824968080267,  
## 42.1989814479909, : some notches went outside hinges ('box'): maybe set  
## notch=FALSE

## Warning in bxp(z = list(stats = structure(c(47.0039971536462,  
## 48.7036350592302, : some notches went outside hinges ('box'): maybe set  
## notch=FALSE

## Warning in bxp(z = list(stats = structure(c(44.8823033479238,  
## 45.5824780104456, : some notches went outside hinges ('box'): maybe set  
## notch=FALSE

## Warning in bxp(z = list(stats = structure(c(40.4351780296797,  
## 41.0055734992057, : some notches went outside hinges ('box'): maybe set  
## notch=FALSE

## Warning in bxp(z = list(stats = structure(c(40.0108052718767,  
## 40.7437102264242, : some notches went outside hinges ('box'): maybe set  
## notch=FALSE

## Warning in bxp(z = list(stats = structure(c(40.0689510090357,  
## 40.8056715348039, : some notches went outside hinges ('box'): maybe set  
## notch=FALSE

## Warning in bxp(z = list(stats = structure(c(42.7765124917976,  
## 43.7809895869652, : some notches went outside hinges ('box'): maybe set  
## notch=FALSE

## Warning in bxp(z = list(stats = structure(c(40.1298272199873,  
## 40.8355536951181, : some notches went outside hinges ('box'): maybe set  
## notch=FALSE

## Warning in bxp(z = list(stats = structure(c(40.1508116932197, 40.887532218988, :  
## some notches went outside hinges ('box'): maybe set notch=FALSE

## Warning in bxp(z = list(stats = structure(c(44.7161509800825,  
## 45.6496508390178, : some notches went outside hinges ('box'): maybe set  
## notch=FALSE

## Warning in bxp(z = list(stats = structure(c(39.9491849110425,  
## 40.6792874441517, : some notches went outside hinges ('box'): maybe set  
## notch=FALSE

## Warning in bxp(z = list(stats = structure(c(40.1593132345114,  
## 40.8960337602797, : some notches went outside hinges ('box'): maybe set  
## notch=FALSE

## Warning in bxp(z = list(stats = structure(c(39.4992397531662,  
## 40.2293422862754, : some notches went outside hinges ('box'): maybe set  
## notch=FALSE

## Warning in bxp(z = list(stats = structure(c(40.2164533369469,  
## 40.9592224102754, : some notches went outside hinges ('box'): maybe set  
## notch=FALSE

## Warning in bxp(z = list(stats = structure(c(40.987519438708, 41.4758088709694, :  
## some notches went outside hinges ('box'): maybe set notch=FALSE

## Warning in bxp(z = list(stats = structure(c(42.3824640945503, 43.537584194164, :  
## some notches went outside hinges ('box'): maybe set notch=FALSE

## Warning in bxp(z = list(stats = structure(c(41.3125250791947,  
## 42.0174435675169, : some notches went outside hinges ('box'): maybe set  
## notch=FALSE

## Warning in bxp(z = list(stats = structure(c(40.1491590953521,  
## 40.8858796211203, : some notches went outside hinges ('box'): maybe set  
## notch=FALSE

## Warning in bxp(z = list(stats = structure(c(44.6113359175811,  
## 45.3673684732142, : some notches went outside hinges ('box'): maybe set  
## notch=FALSE

## Warning in bxp(z = list(stats = structure(c(40.1366211133071,  
## 40.8667236464162, : some notches went outside hinges ('box'): maybe set  
## notch=FALSE

## Warning in bxp(z = list(stats = structure(c(40.1161326115834,  
## 40.8462351446925, : some notches went outside hinges ('box'): maybe set  
## notch=FALSE

## Warning in bxp(z = list(stats = structure(c(40.4483573603222,  
## 40.9989361112592, : some notches went outside hinges ('box'): maybe set  
## notch=FALSE

## Warning in bxp(z = list(stats = structure(c(42.9004521600794,  
## 43.7301942062136, : some notches went outside hinges ('box'): maybe set  
## notch=FALSE

## Warning in bxp(z = list(stats = structure(c(39.8008660651757,  
## 40.4877727273369, : some notches went outside hinges ('box'): maybe set  
## notch=FALSE

## Warning in bxp(z = list(stats = structure(c(41.0260182197014,  
## 42.0320149859846, : some notches went outside hinges ('box'): maybe set  
## notch=FALSE

## Warning in bxp(z = list(stats = structure(c(41.2584720444697,  
## 41.8966760297647, : some notches went outside hinges ('box'): maybe set  
## notch=FALSE

## Warning in bxp(z = list(stats = structure(c(42.7092516181002,  
## 43.5342237369968, : some notches went outside hinges ('box'): maybe set  
## notch=FALSE

## Warning in bxp(z = list(stats = structure(c(40.1728926416597, 40.909613167428, :  
## some notches went outside hinges ('box'): maybe set notch=FALSE

## Warning in bxp(z = list(stats = structure(c(40.029314529585, 40.7660350553532, :  
## some notches went outside hinges ('box'): maybe set notch=FALSE

## Warning in bxp(z = list(stats = structure(c(40.2737737863724,  
## 41.0174013632256, : some notches went outside hinges ('box'): maybe set  
## notch=FALSE

## Warning in bxp(z = list(stats = structure(c(43.9092211885852,  
## 45.1020987332039, : some notches went outside hinges ('box'): maybe set  
## notch=FALSE

## Warning in bxp(z = list(stats = structure(c(41.8926540723089,  
## 42.8856483062973, : some notches went outside hinges ('box'): maybe set  
## notch=FALSE

## Warning in bxp(z = list(stats = structure(c(44.852225603291, 46.0810854751204, :  
## some notches went outside hinges ('box'): maybe set notch=FALSE

## Warning in bxp(z = list(stats = structure(c(40.1491590953521,  
## 40.8858796211203, : some notches went outside hinges ('box'): maybe set  
## notch=FALSE

## Warning in bxp(z = list(stats = structure(c(41.6913118632307,  
## 42.2897433280924, : some notches went outside hinges ('box'): maybe set  
## notch=FALSE

## Warning in bxp(z = list(stats = structure(c(40.1421459612021,  
## 40.8722484943113, : some notches went outside hinges ('box'): maybe set  
## notch=FALSE

## Warning in bxp(z = list(stats = structure(c(40.0789701555178, 40.761378048367, :  
## some notches went outside hinges ('box'): maybe set notch=FALSE

## Warning in bxp(z = list(stats = structure(c(40.5913376655775,  
## 41.2632527375694, : some notches went outside hinges ('box'): maybe set  
## notch=FALSE

## Warning in bxp(z = list(stats = structure(c(42.6971022111099,  
## 43.8178853760988, : some notches went outside hinges ('box'): maybe set  
## notch=FALSE

pdf(file = "v2",   
 width = 5,   
 height = 5)   
plot.variable(grow, 'V2', partial = TRUE, notch = FALSE)  
  
pdf(file = "v3",   
 width = 5,   
 height = 5)   
plot.variable(grow, 'V3', partial = TRUE, notch = FALSE)  
  
pdf(file = "v4",   
 width = 5,   
 height = 5)   
plot.variable(grow, 'V4', partial = TRUE, notch = FALSE)  
  
pdf(file = "v5",   
 width = 5,   
 height = 5)   
plot.variable(grow, 'V5', partial = TRUE, notch = FALSE)  
  
pdf(file = "v6",   
 width = 5,   
 height = 5)   
plot.variable(grow, 'V6', partial = TRUE, notch = FALSE)  
  
pdf(file = "v7",   
 width = 5,   
 height = 5)   
plot.variable(grow, 'V7', partial = TRUE, notch = FALSE)  
  
pdf(file = "v8",   
 width = 5,   
 height = 5)   
plot.variable(grow, 'V8', partial = TRUE, notch = FALSE)  
  
pdf(file = "v9",   
 width = 5,   
 height = 5)   
plot.variable(grow, 'V9', partial = TRUE, notch = FALSE)  
  
pdf(file = "v10",   
 width = 5,   
 height = 5)   
plot.variable(grow, 'V10', partial = TRUE, notch = FALSE)  
  
 dev.off()

## png   
## 2

################arima#############  
  
  
weeklyseveritygermany <- read\_excel("Downloads/weeklyseveritygermany(1).xlsx")

## New names:  
## \* `` -> ...5

data <- weeklyseveritygermany[6:74,] #skip the first 6 to make it withthe same length with RF embd-  
datatrain<-data[1:53,]  
datatest<-data[54:68,] #from 2021 02 14 to 05.09  
  
xregtrain<-as.matrix(datatrain[c(12, 14, 16, 18, 20, 22, 24, 26, 47)])  
xregtest<-as.matrix(datatest[c(12, 14, 16, 18, 20, 22, 24, 26, 47)])  
  
  
##### here it works ########  
library(forecast)  
library(fable)

## Loading required package: fabletools

##   
## Attaching package: 'fabletools'

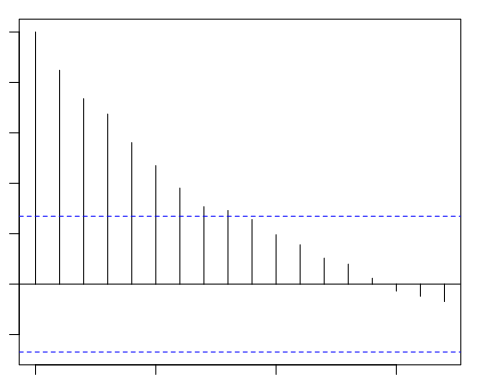
## The following object is masked \_by\_ '.GlobalEnv':  
##   
## RMSE

## The following objects are masked from 'package:forecast':  
##   
## accuracy, forecast

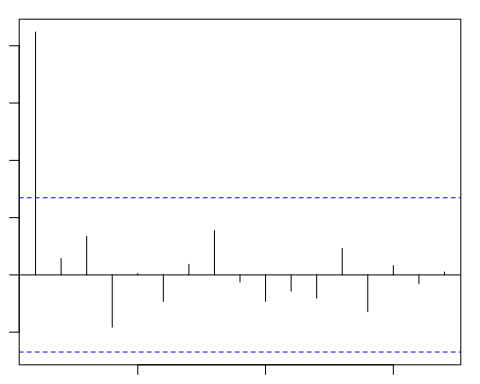
library(feasts)  
ndiffs(datatrain$KH)

## [1] 2

acf(datatrain$KH)



pacf(datatrain$KH)



arimatrain <- Arima(datatrain$KH, order = c(1, 2, 1),  
 xreg = xregtrain, lambda=BoxCox.lambda(datatrain$KH))  
BoxCox.lambda(datatrain$KH)

## [1] 0.3177898

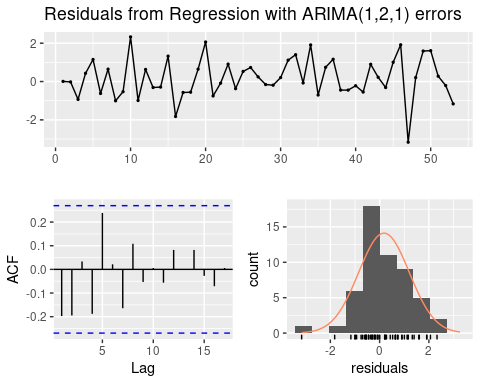
residuals(arimatrain)

## Time Series:  
## Start = 1   
## End = 53   
## Frequency = 1   
## [1] 0.006612283 -0.019574655 -0.933637844 0.430834576 1.150581320  
## [6] -0.623350309 0.646053257 -1.005420444 -0.530280928 2.330241510  
## [11] -0.988086104 0.630718852 -0.311402936 -0.290248805 1.326110563  
## [16] -1.823973695 -0.571933605 -0.557149456 0.645215847 2.063009361  
## [21] -0.751516402 -0.084960506 0.910341097 -0.372569945 0.528669022  
## [26] 0.732396295 0.243388065 -0.156862302 -0.190969794 0.205223366  
## [31] 1.114978174 1.399997667 -0.075209238 1.910075786 -0.702947706  
## [36] 0.744861871 1.161921864 -0.446498079 -0.446899411 -0.220108491  
## [41] -0.549525543 0.900182766 0.222069664 -0.313082869 1.004450787  
## [46] 1.917882374 -3.165642471 0.206089837 1.591227327 1.611170773  
## [51] 0.279064682 -0.208956853 -1.166296879

coef(arimatrain)

## ar1 ma1 school workplace   
## -0.5447162 -0.9017151 0.0304845 -1.2681981   
## publicevent gatherings publictransport stayathome   
## -0.1134983 0.1608419 -0.5846091 -0.3444371   
## internalmovement internationaltravel StringencyIndex   
## 0.2831173 0.8175627 -0.0299140

checkresiduals(arimatrain)

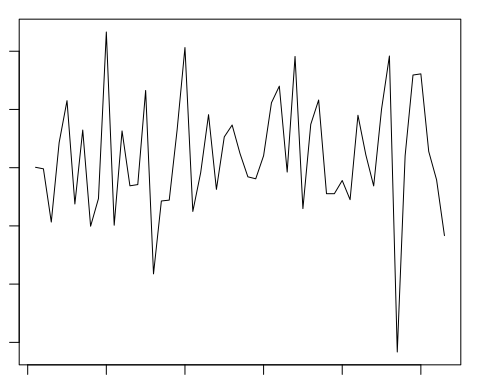


##   
## Ljung-Box test  
##   
## data: Residuals from Regression with ARIMA(1,2,1) errors  
## Q\* = 13.863, df = 3, p-value = 0.003098  
##   
## Model df: 11. Total lags used: 14

arimatrain

## Series: datatrain$KH   
## Regression with ARIMA(1,2,1) errors   
## Box Cox transformation: lambda= 0.3177898   
##   
## Coefficients:  
## ar1 ma1 school workplace publicevent gatherings  
## -0.5447 -0.9017 0.0305 -1.2682 -0.1135 0.1608  
## s.e. 0.1274 0.0565 0.4580 1.0411 0.9805 0.5088  
## publictransport stayathome internalmovement internationaltravel  
## -0.5846 -0.3444 0.2831 0.8176  
## s.e. 1.2431 0.4950 0.5312 0.7615  
## StringencyIndex  
## -0.0299  
## s.e. 0.1156  
##   
## sigma^2 estimated as 1.418: log likelihood=-76.48  
## AIC=176.96 AICc=185.18 BIC=200.15

plot(residuals(arimatrain))



fc<-forecast(arimatrain, as.ts(datatest$KH), xreg=xregtest, lambda=BoxCox.lambda(datatrain$KH) )

## Warning in InvBoxCox(pred$pred, lambda, biasadj, pred$se^2): biasadj information  
## not found, defaulting to FALSE.

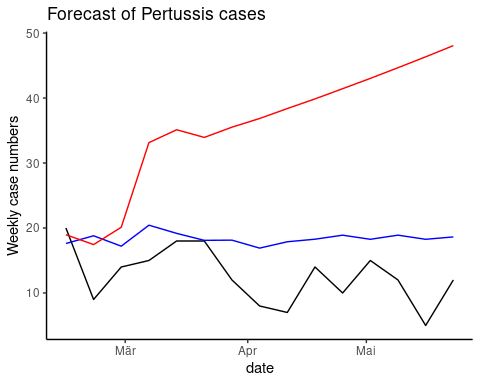
accuracy(fc, datatest$KH)

## ME RMSE MAE MPE MAPE MASE  
## Training set 1.544017 13.41073 9.057179 -5.294251 34.08395 0.7771836  
## Test set -22.924683 25.65886 23.062736 -235.159957 235.85022 1.9789807  
## ACF1  
## Training set -0.3530218  
## Test set NA

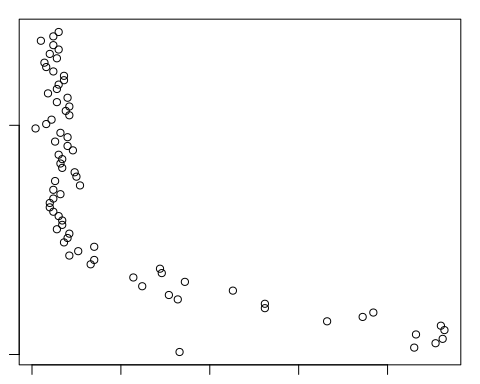
y\_pred <- ts(pred$predicted  
 ,  
 start = c(2021, 1),   
 frequency = 52  
)  
# add the forecasts to the original tibble  
datatest$predrf<-y\_pred  
datatest$predarima<-fc$mean  
dm.test(datatest$KH-datatest$predrf,datatest$KH-datatest$predarima )

##   
## Diebold-Mariano Test  
##   
## data: datatest$KH - datatest$predrfdatatest$KH - datatest$predarima  
## DM = -5.0079, Forecast horizon = 1, Loss function power = 2, p-value =  
## 0.0001917  
## alternative hypothesis: two.sided

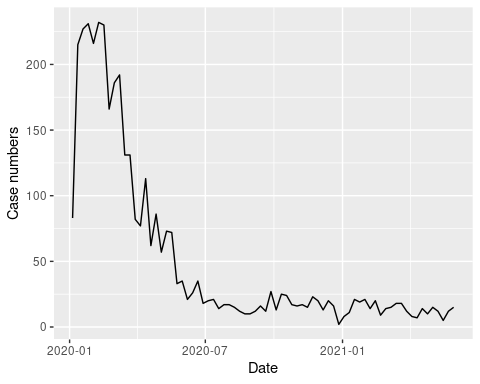
# visualize the forecasts  
plot\_fc <- datatest %>%   
 ggplot(aes(x = date)) +  
 geom\_line(aes(y = KH)) +  
 geom\_line(aes(y = predrf), color = "blue") +  
 geom\_line(aes(y = predarima), color="red") +  
 theme\_classic() +  
 labs(  
 title = "Forecast of Pertussis cases",  
 y = "Weekly case numbers"  
 )  
plot\_fc



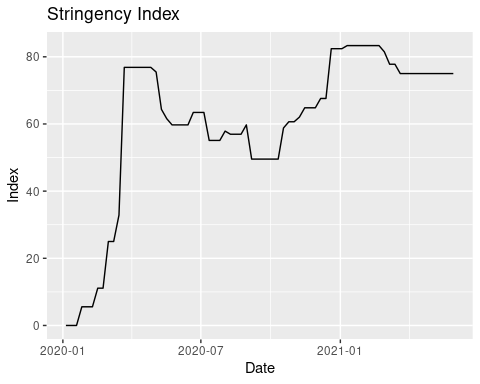
###### plots########  
  
   
  
library(ggplot2)  
  
   
  
plot(weeklyseveritygermany$KH, weeklyseveritygermany$date)



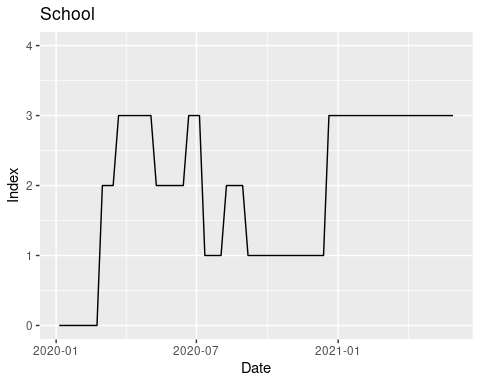
datagraph <- ggplot(weeklyseveritygermany, aes(x = date, y = KH)) +  
  
 geom\_line()+  
  
 labs (y= "Case numbers", x = "Date")  
  
datagraph



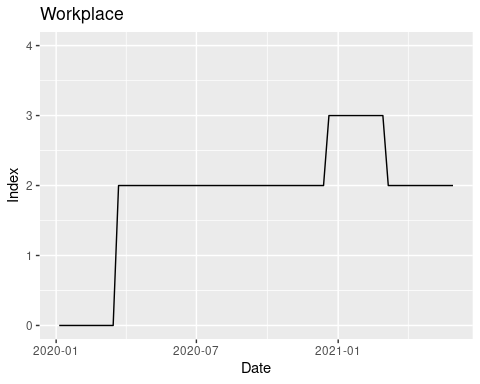
predictor\_si <- ggplot(weeklyseveritygermany, aes(x = date, y = StringencyIndex)) +  
  
 geom\_line()+  
  
 expand\_limits(y=c(0, 4)) +  
  
 labs (y= "Index", x = "Date")+  
  
 ggtitle ("Stringency Index")  
  
predictor\_si



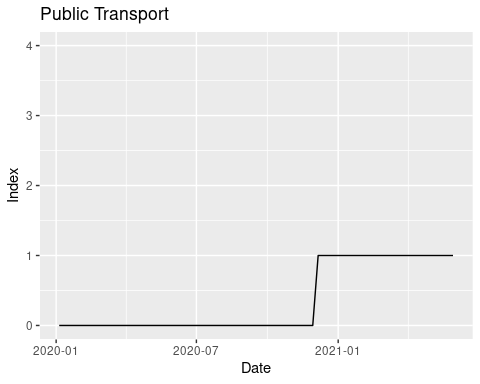
predictor\_school <- ggplot(weeklyseveritygermany, aes(x = date, y = school)) +  
  
 geom\_line()+  
  
 expand\_limits(y=c(0, 4)) +  
  
 labs (y= "Index", x = "Date")+  
  
ggtitle ("School")  
  
predictor\_school



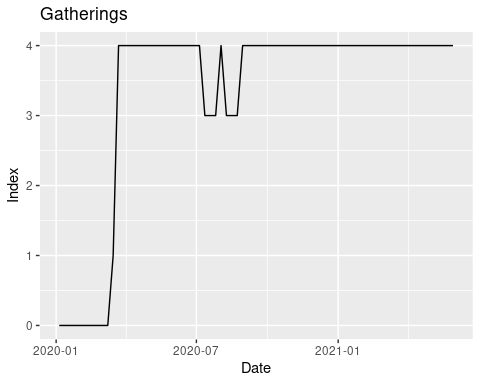
predictor\_work <- ggplot(weeklyseveritygermany, aes(x = date, y = workplace)) +  
  
 geom\_line()+  
  
 expand\_limits(y=c(0, 4)) +  
  
 labs (y= "Index", x = "Date") +  
  
 ggtitle ("Workplace")  
  
predictor\_work



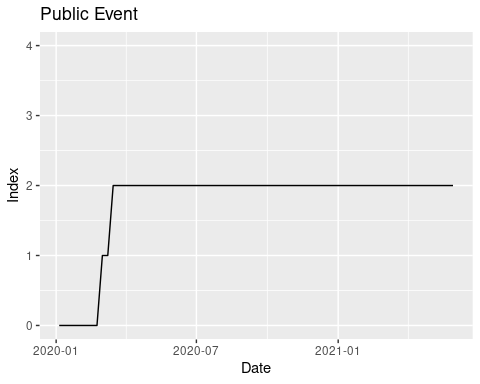
predictor\_publictransport <- ggplot(weeklyseveritygermany, aes(x = date, y = publictransport)) +  
  
 geom\_line()+  
  
 expand\_limits(y=c(0, 4)) +  
  
 labs (y= "Index", x = "Date") +  
  
 ggtitle("Public Transport")  
  
predictor\_publictransport



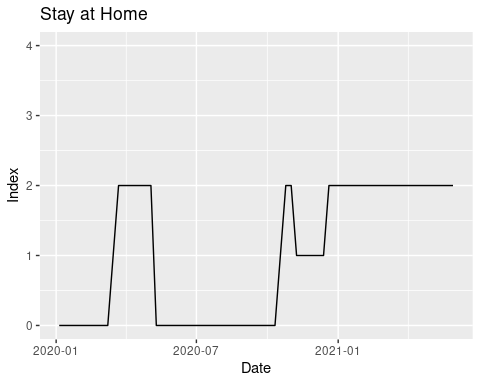
predictor\_gathering <- ggplot(weeklyseveritygermany, aes(x = date, y = gatherings)) +  
  
 geom\_line()+  
  
 expand\_limits(y=c(0, 4)) +  
  
 labs (y= "Index", x = "Date") +  
  
 ggtitle("Gatherings")  
  
predictor\_gathering



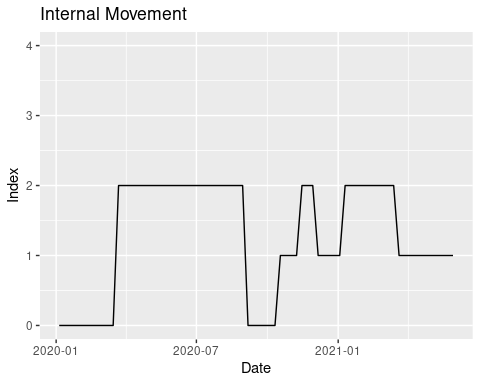
predictor\_publicevent <- ggplot(weeklyseveritygermany, aes(x = date, y = publicevent)) +  
  
 geom\_line()+  
  
 expand\_limits(y=c(0, 4)) +  
  
 labs (y= "Index", x = "Date") +  
  
 ggtitle("Public Event")  
  
predictor\_publicevent



predictor\_stayathome <- ggplot(weeklyseveritygermany, aes(x = date, y = stayathome)) +  
  
 geom\_line()+  
  
 expand\_limits(y=c(0, 4)) +  
  
 labs (y= "Index", x = "Date")+  
  
 ggtitle("Stay at Home")  
  
predictor\_stayathome



predictor\_international <- ggplot(weeklyseveritygermany, aes(x = date, y = internationatravel)) +  
  
 geom\_line()+  
  
 expand\_limits(y=c(0, 4)) +  
  
 labs (y= "Index", x = "Date")  
#predictor\_international  
  
  
   
  
predictor\_internal <- ggplot(weeklyseveritygermany, aes(x = date, y = internalmovement)) +  
  
 geom\_line()+  
  
 expand\_limits(y=c(0, 4)) +  
  
 labs (y= "Index", x = "Date") +  
  
 ggtitle("Internal Movement")  
  
predictor\_internal



sessionInfo()

## R version 4.0.5 (2021-03-31)  
## Platform: x86\_64-pc-linux-gnu (64-bit)  
## Running under: Ubuntu 20.04.2 LTS  
##   
## Matrix products: default  
## BLAS: /usr/lib/x86\_64-linux-gnu/blas/libblas.so.3.9.0  
## LAPACK: /usr/lib/x86\_64-linux-gnu/lapack/liblapack.so.3.9.0  
##   
## locale:  
## [1] LC\_CTYPE=en\_US.UTF-8 LC\_NUMERIC=C   
## [3] LC\_TIME=de\_DE.UTF-8 LC\_COLLATE=en\_US.UTF-8   
## [5] LC\_MONETARY=de\_DE.UTF-8 LC\_MESSAGES=en\_US.UTF-8   
## [7] LC\_PAPER=de\_DE.UTF-8 LC\_NAME=C   
## [9] LC\_ADDRESS=C LC\_TELEPHONE=C   
## [11] LC\_MEASUREMENT=de\_DE.UTF-8 LC\_IDENTIFICATION=C   
##   
## attached base packages:  
## [1] stats graphics grDevices utils datasets methods base   
##   
## other attached packages:  
## [1] feasts\_0.1.7 fable\_0.3.0 fabletools\_0.3.0   
## [4] ggRandomForests\_2.0.1 randomForestSRC\_2.11.0 forecast\_8.14   
## [7] randomForest\_4.6-14 tsibble\_1.0.0 forcats\_0.5.1   
## [10] stringr\_1.4.0 dplyr\_1.0.6 purrr\_0.3.4   
## [13] readr\_1.4.0 tidyr\_1.1.3 tibble\_3.1.2   
## [16] ggplot2\_3.3.3 tidyverse\_1.3.1 readxl\_1.3.1   
##   
## loaded via a namespace (and not attached):  
## [1] nlme\_3.1-152 fs\_1.5.0 xts\_0.12.1   
## [4] lubridate\_1.7.10 RColorBrewer\_1.1-2 httr\_1.4.2   
## [7] data.tree\_1.0.0 tools\_4.0.5 backports\_1.2.1   
## [10] utf8\_1.1.4 R6\_2.5.0 DBI\_1.1.1   
## [13] colorspace\_2.0-0 nnet\_7.3-15 withr\_2.3.0   
## [16] tidyselect\_1.1.0 curl\_4.3 compiler\_4.0.5   
## [19] cli\_2.5.0 rvest\_1.0.0 xml2\_1.3.2   
## [22] labeling\_0.4.2 tseries\_0.10-48 scales\_1.1.1   
## [25] lmtest\_0.9-38 fracdiff\_1.5-1 quadprog\_1.5-8   
## [28] digest\_0.6.27 rmarkdown\_2.7 pkgconfig\_2.0.3   
## [31] htmltools\_0.5.1.1 dbplyr\_2.1.1 htmlwidgets\_1.5.3   
## [34] rlang\_0.4.10 TTR\_0.24.2 rstudioapi\_0.13   
## [37] quantmod\_0.4.18 visNetwork\_2.0.9 generics\_0.1.0   
## [40] farver\_2.0.3 zoo\_1.8-8 jsonlite\_1.7.2   
## [43] distributional\_0.2.2 magrittr\_2.0.1 Matrix\_1.3-2   
## [46] Rcpp\_1.0.6 munsell\_0.5.0 fansi\_0.4.1   
## [49] lifecycle\_1.0.0 stringi\_1.5.3 yaml\_2.2.1   
## [52] grid\_4.0.5 parallel\_4.0.5 crayon\_1.4.1   
## [55] lattice\_0.20-41 haven\_2.3.1 splines\_4.0.5   
## [58] hms\_1.0.0 knitr\_1.31 anytime\_0.3.9   
## [61] pillar\_1.6.1 reprex\_2.0.0 urca\_1.3-0   
## [64] glue\_1.4.2 evaluate\_0.14 modelr\_0.1.8   
## [67] vctrs\_0.3.8 cellranger\_1.1.0 gtable\_0.3.0   
## [70] assertthat\_0.2.1 xfun\_0.20 broom\_0.7.6   
## [73] survival\_3.2-10 timeDate\_3043.102 DiagrammeR\_1.0.6.1   
## [76] ellipsis\_0.3.2