assignment\_5

Harinath Reddy

2022-11-16

library(cdcfluview); library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(tsibble); library(lubridate)

##   
## Attaching package: 'tsibble'

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, union

## Loading required package: timechange

##   
## Attaching package: 'lubridate'

## The following object is masked from 'package:tsibble':  
##   
## interval

## The following objects are masked from 'package:base':  
##   
## date, intersect, setdiff, union

library(fable);library(ggplot2)

## Loading required package: fabletools

library(feasts)

# Prepare the data  
usflu.raw <- ilinet("national", years = 2010:2018)  
names(usflu.raw)

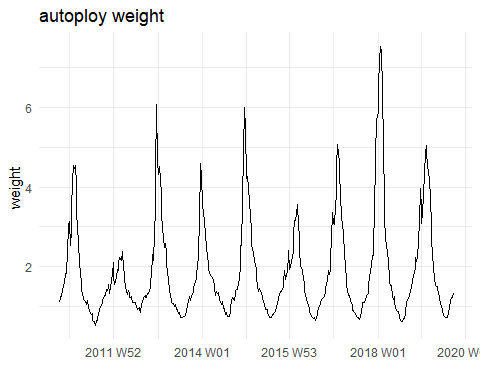
## [1] "region\_type" "region" "year" "week"   
## [5] "weighted\_ili" "unweighted\_ili" "age\_0\_4" "age\_25\_49"   
## [9] "age\_25\_64" "age\_5\_24" "age\_50\_64" "age\_65"   
## [13] "ilitotal" "num\_of\_providers" "total\_patients" "week\_start"

usflu <- usflu.raw %>%  
 mutate(  
 date = as.Date(paste0(year, sprintf("%02d", week), "00"),  
format="%Y%W%w"),  
dec\_date = decimal\_date(week\_start),  
week = yearweek(week\_start),  
time\_in\_year = dec\_date%%1)%>%  
 dplyr::filter(!is.na(dec\_date))

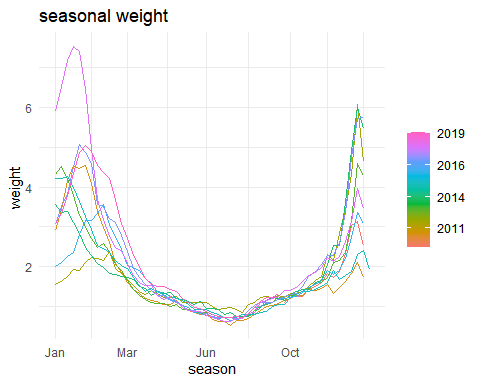
## Warning in strptime(x, format, tz = "GMT"): (0-based) yday 368 in year 2014 is  
## invalid

usflu.ts <- as\_tsibble(usflu, index = week)

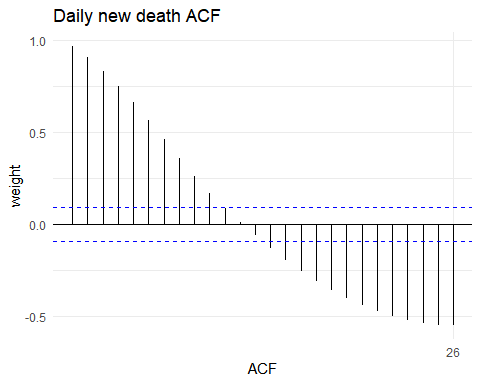
autoplot.weighted\_ili<- usflu.ts %>% autoplot(weighted\_ili) + theme\_minimal() +   
 labs(title="autoploy weight",  
 x="", y="weight ")  
autoplot.weighted\_ili



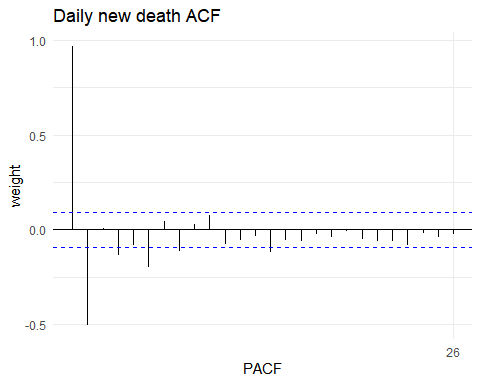
gg\_season.weight <- usflu.ts %>% gg\_season(weighted\_ili) + theme\_minimal() +   
 labs(title="seasonal weight ", x="season", y="weight ")  
gg\_season.weight



acf.weight<- usflu.ts %>% ACF(weighted\_ili) %>%  
 autoplot() + theme\_minimal() + labs(title="Daily new death ACF",  
 x="ACF", y="weight")  
acf.weight

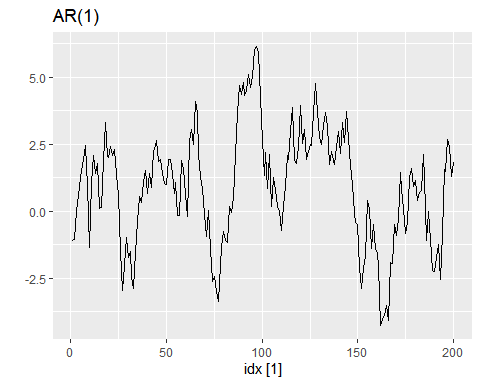


pacf.weight<- usflu.ts %>% PACF(weighted\_ili) %>%  
 autoplot() + theme\_minimal() + labs(title="Daily new death ACF",  
 x="PACF", y="weight")  
pacf.weight



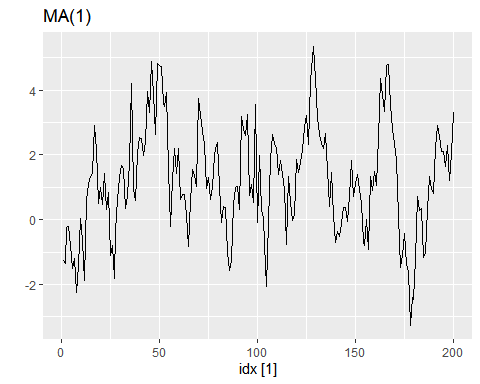
\*\* (a) AR(1) model with ϕ1 = 0.9 and σ2 = 1.

tsibble(idx = seq\_len(200), sim = 1 + arima.sim(list(ar = c(0.9)), n = 200), index = idx) %>%  
autoplot(sim) + ylab("") + ggtitle("AR(1)")

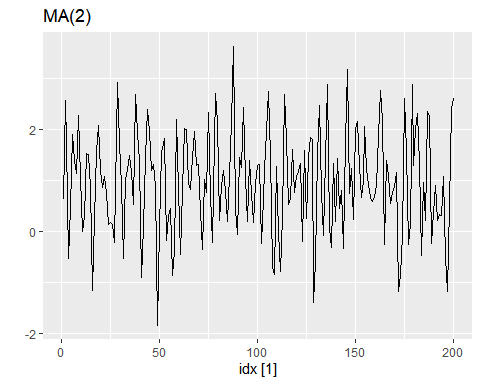


\*\* (a) AR(1) model with ϕ1 = 0.9 and σ2 = 1.

tsibble(idx = seq\_len(200), sim = 1 + arima.sim(list(ar = c(0.8)), n = 200), index = idx) %>%  
autoplot(sim) + ylab("") + ggtitle("MA(1)")



tsibble(idx = seq\_len(200), sim = 1 + arima.sim(list(ar = c(0.3,-0.4)), n = 200), index = idx) %>%  
autoplot(sim) + ylab("") + ggtitle("MA(2)")



library(IDDA)  
data("state.long")

va\_state<- IDDA::state.long%>%  
 filter(State=="Virginia")  
  
va\_state

## # A tibble: 345 × 7  
## State Region Division pop DATE Infected Death  
## <chr> <fct> <fct> <int> <date> <int> <int>  
## 1 Virginia South South Atlantic 8517685 2020-12-31 349584 5032  
## 2 Virginia South South Atlantic 8517685 2020-12-30 344343 4982  
## 3 Virginia South South Atlantic 8517685 2020-12-29 340297 4918  
## 4 Virginia South South Atlantic 8517685 2020-12-28 336173 4857  
## 5 Virginia South South Atlantic 8517685 2020-12-27 333570 4850  
## 6 Virginia South South Atlantic 8517685 2020-12-26 329575 4833  
## 7 Virginia South South Atlantic 8517685 2020-12-25 327990 4816  
## 8 Virginia South South Atlantic 8517685 2020-12-24 323913 4788  
## 9 Virginia South South Atlantic 8517685 2020-12-23 319131 4757  
## 10 Virginia South South Atlantic 8517685 2020-12-22 314479 4701  
## # … with 335 more rows

\*\*\* splitting the data into training and testing sets also seperating data from dependent variable\*\*\*

state.ts <- as\_tsibble(state.long, key = State) %>%  
group\_by(State) %>%  
mutate(Infected = Infected/1000) %>%  
mutate(YDA\_Infected = lag(Infected, order\_by = DATE)) %>%  
mutate(YDA\_Death = lag(Death, order\_by = DATE)) %>%  
mutate(Y.Infected = Infected - YDA\_Infected) %>%  
mutate(Y.Death = Death - YDA\_Death) %>%  
mutate(cum\_infected = cumsum(Infected))%>%  
mutate(cum\_death = cumsum(Death)) %>%  
dplyr::filter(!is.na(Y.Infected)) %>%  
dplyr::filter(!is.na(Y.Death)) %>%  
dplyr::select(-c(YDA\_Infected, YDA\_Death))%>%  
filter(State=="Virginia")

## Using `DATE` as index variable.

Virginia.ts <- state.ts %>%  
dplyr::filter(State == "Virginia") %>%  
dplyr::select(Infected, Death, cum\_infected, cum\_death, Y.Death, Y.Infected)

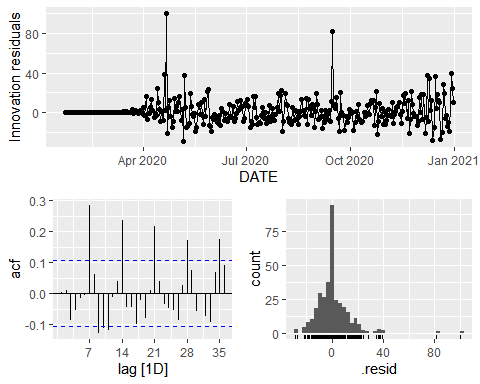
## Adding missing grouping variables: `State`

arima.fit <-Virginia.ts %>%  
 model(arima = ARIMA(Y.Death ~ PDQ(0,0,0)))  
report(arima.fit)

## Series: Y.Death   
## Model: ARIMA(0,1,2)   
##   
## Coefficients:  
## ma1 ma2  
## -0.6529 -0.2045  
## s.e. 0.0510 0.0500  
##   
## sigma^2 estimated as 164.6: log likelihood=-1361.58  
## AIC=2729.15 AICc=2729.22 BIC=2740.67

## Selected Model is ARIMA(0,1,2) with Log-likelyhood estimator of -1361.58 which is infact very poor.

arima.fit %>% gg\_tsresiduals(lag=36)



## The resudials plot seems obvious that the error terms are normally distributed.