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)
```

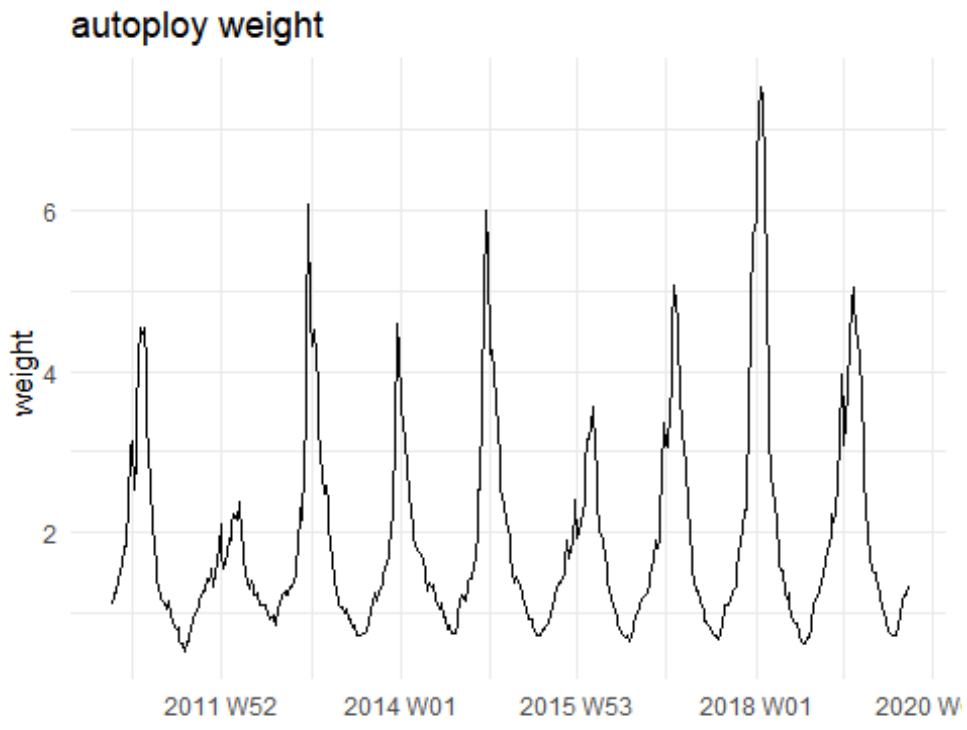
```
##  [1] "region_type"      "region"           "year"             "week"
##  [5] "weighted_ili"     "unweighted_ili"   "age_0_4"          "age_25_49"
##  [9] "age_25_64"        "age_5_24"         "age_50_64"        "age_65"
## [13] "ilitotal"         "num_of_providers" "total_patients"   "week_start"
```

```
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))
```

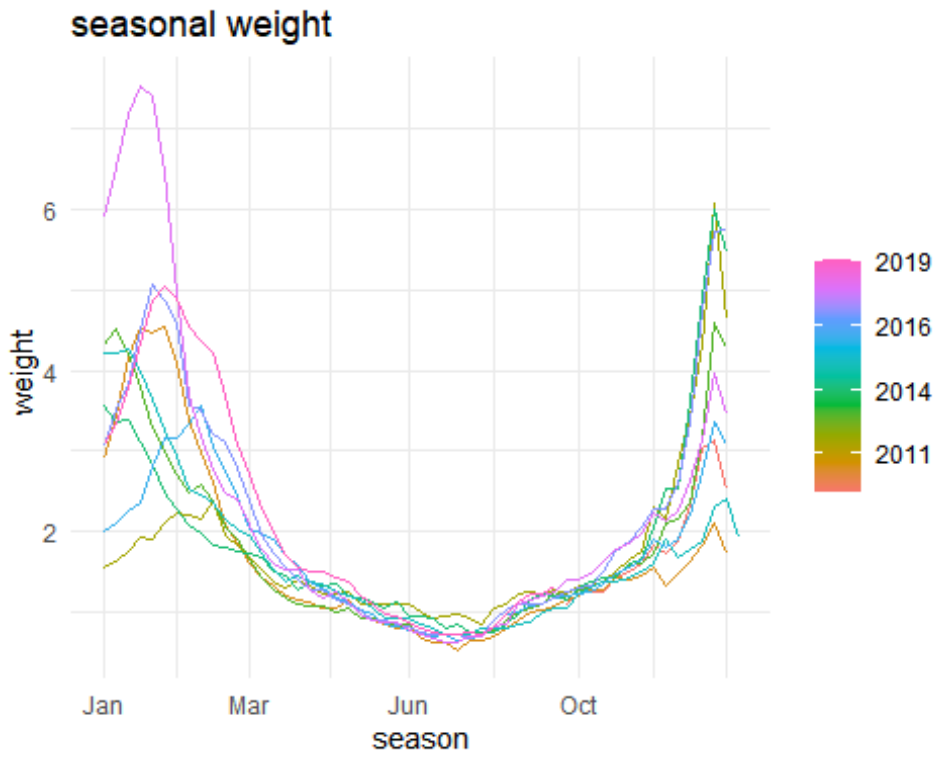
```
## Warning in strptime(x, format, tz = "GMT"): (0-based) yday 368 in year
2014 is
## invalid
```

```
usflu.ts <- as_tsibble(usflu, index = week)
```

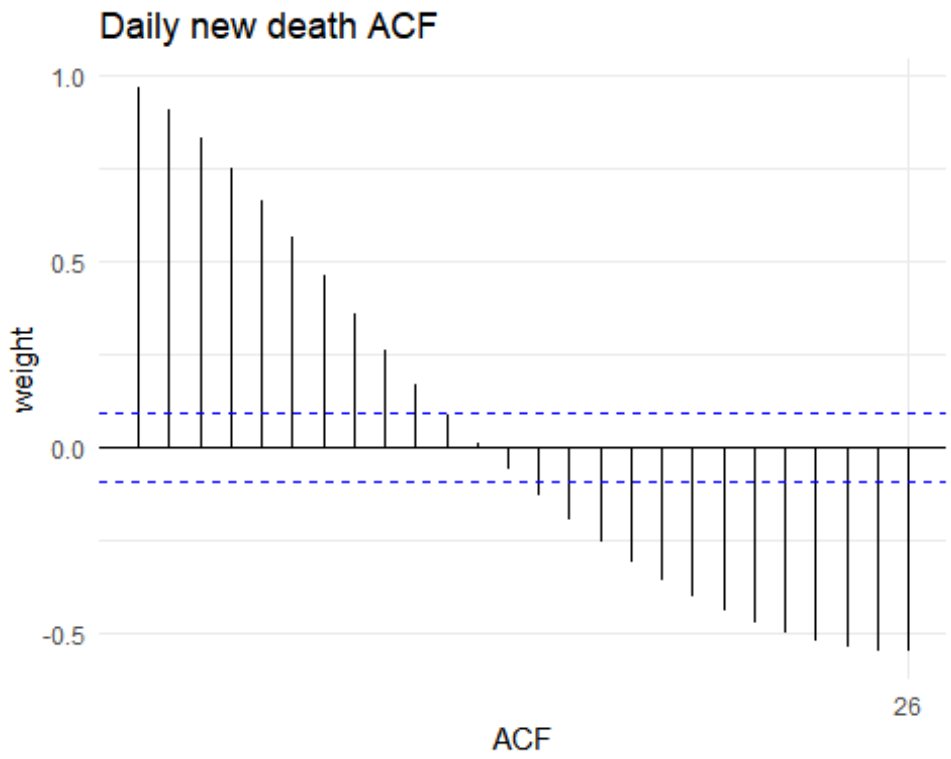
```
autoplot.weighted_ili<- usflu.ts %>% autoplot(weighted_ili) + theme_minimal()
+
  labs(title="autoploy weight",
       x="", y="weight ")
autoplot.weighted_ili
```



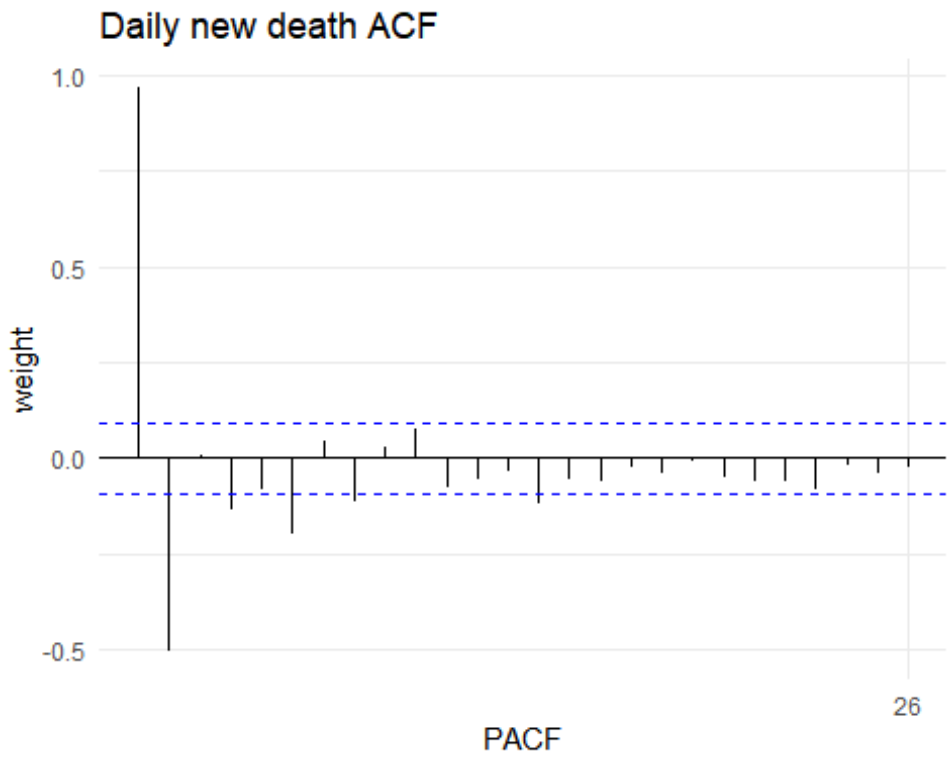
```
gg_season.weight <- usflu.ts %>% gg_season(weighted_ili) + theme_minimal() +
  labs(title="seasonal weight ", x="season", y="weight ")
gg_season.weight
```




```
acf.weight<- usflu.ts %>% ACF(weighted_ili) %>%
  autoplot() + theme_minimal() + labs(title="Daily new death ACF",
                          x="ACF", y="weight")
acf.weight
```

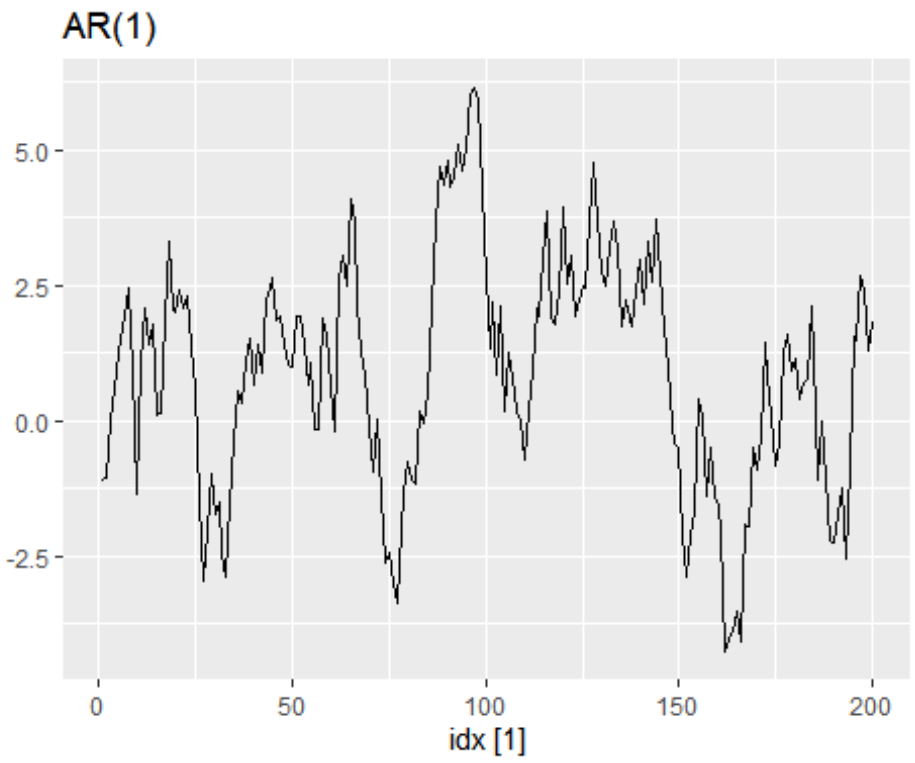



```
pacf.weight<- usflu.ts %>% PACF(weighted_ili) %>%
  autoplot() + theme_minimal() + labs(title="Daily new death ACF",
                          x="PACF", y="weight")
pacf.weight
```



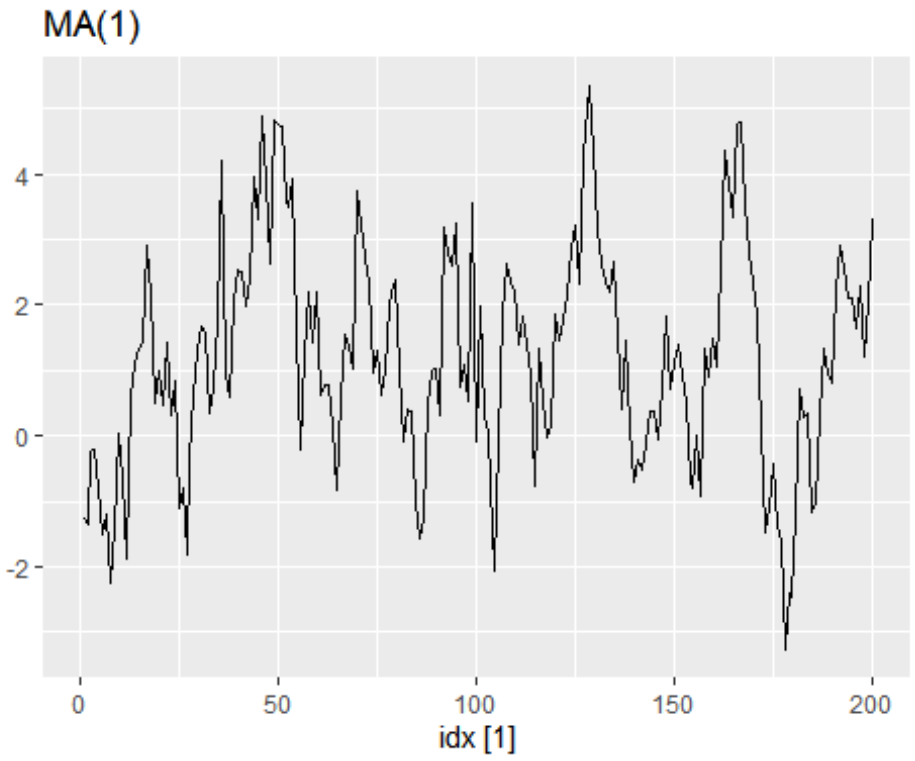
** (a) AR(1) model with ϕ1 = 0.9 and σ2 = 1.

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

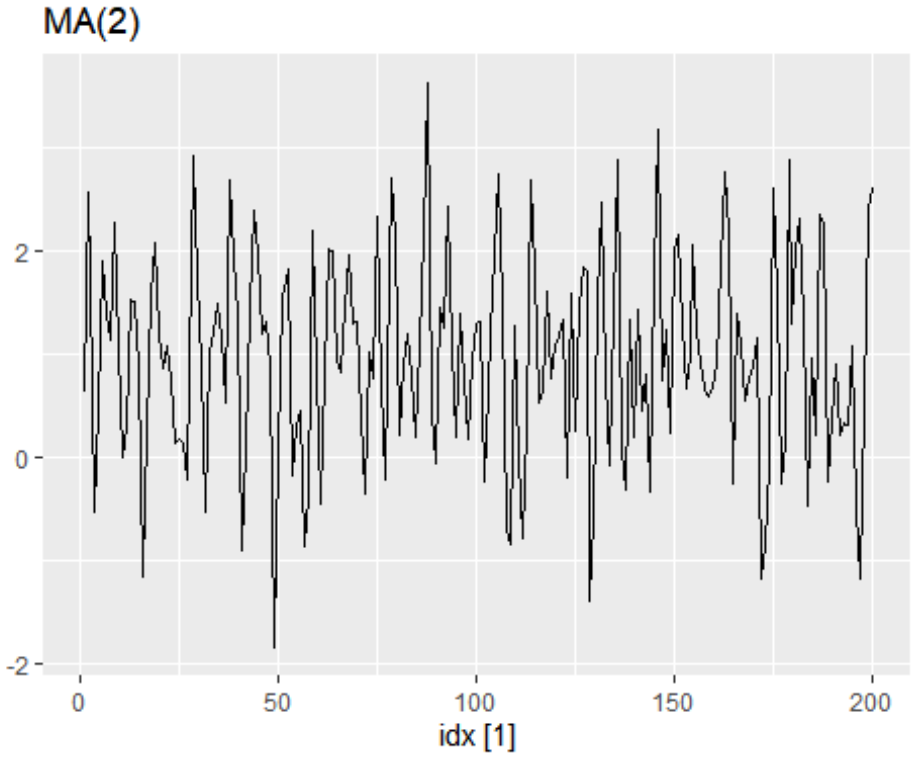


** (a) AR(1) model with ϕ1 = 0.9 and σ2 = 1.

```
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)
```

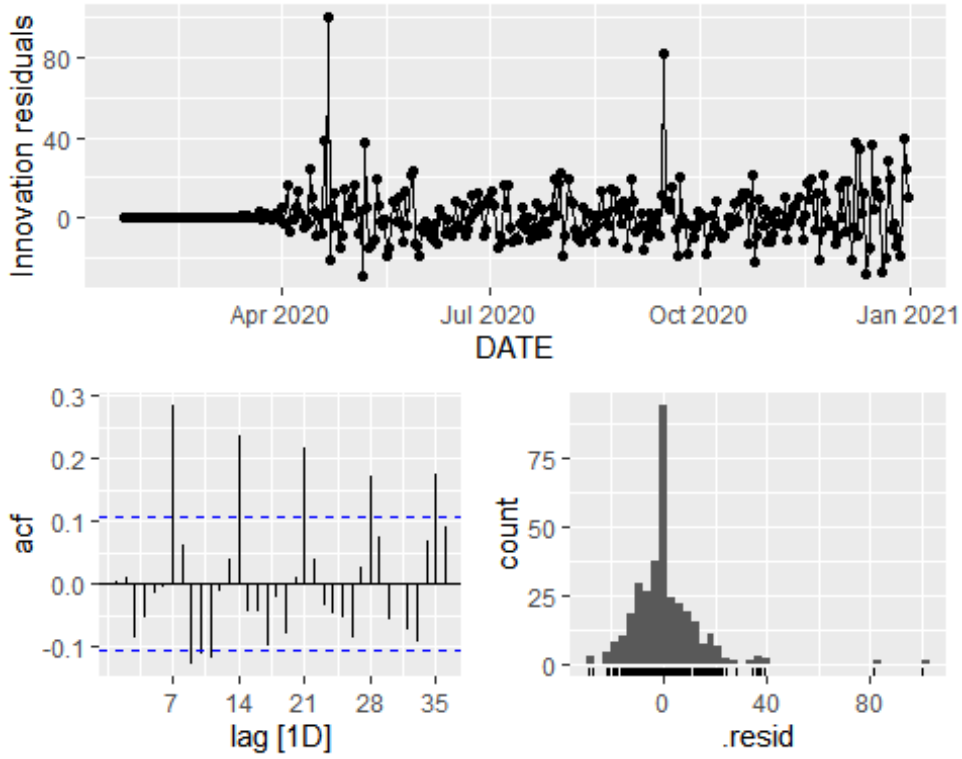
```
## Adding missing grouping variables: `State`
```

```
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.
```