Covid19

Julien JACQUES

The data are available at https://covid19.who.int/WHO-COVID-19-global-data.csv

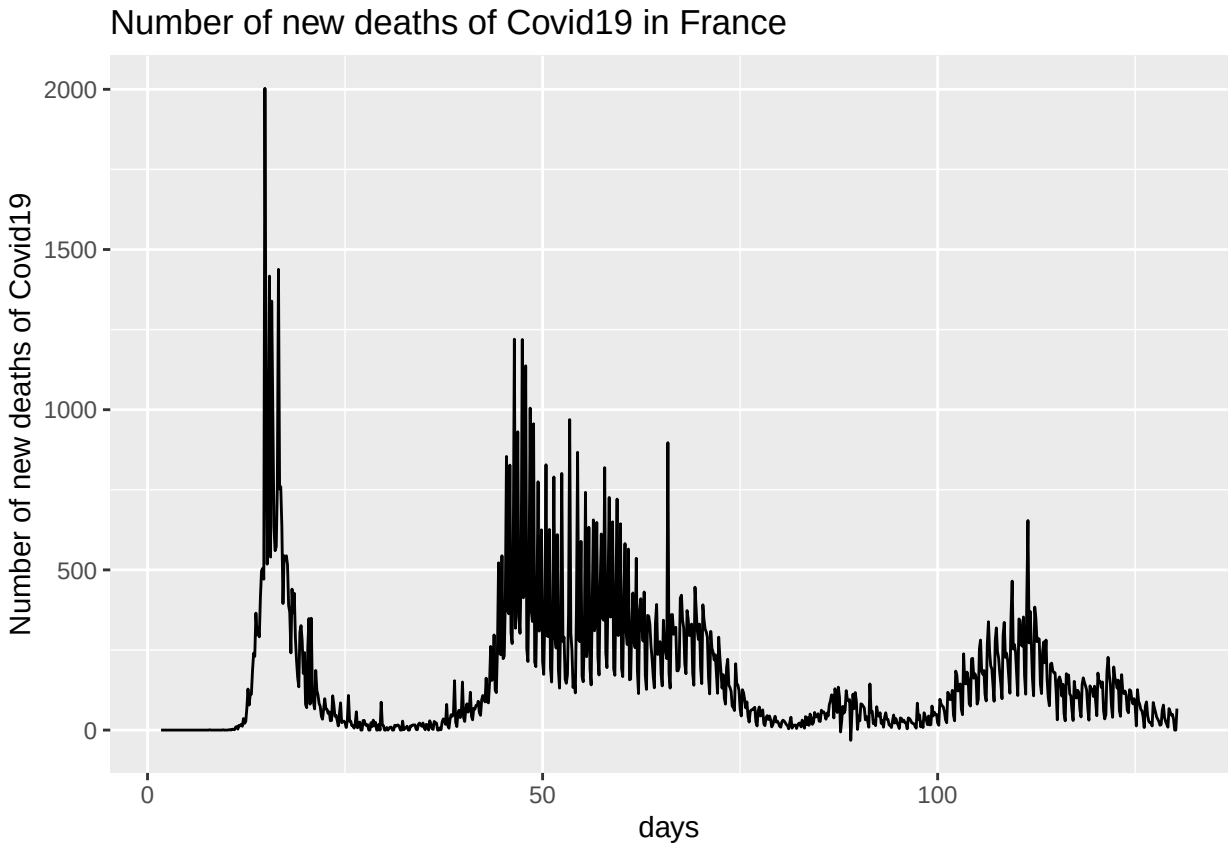
```
covid19 = read.csv("data/2022-06-22-WHO-COVID-19-global-data.csv")
covid19_F=covid19[covid19$Country=="France",]
covid19_F_nc=ts(covid19_F$New_cases,freq=7,start=c(1,6))
covid19_F_nd=ts(covid19_F$New_deaths,freq=7,start=c(1,6))
```

We plot the number of new death

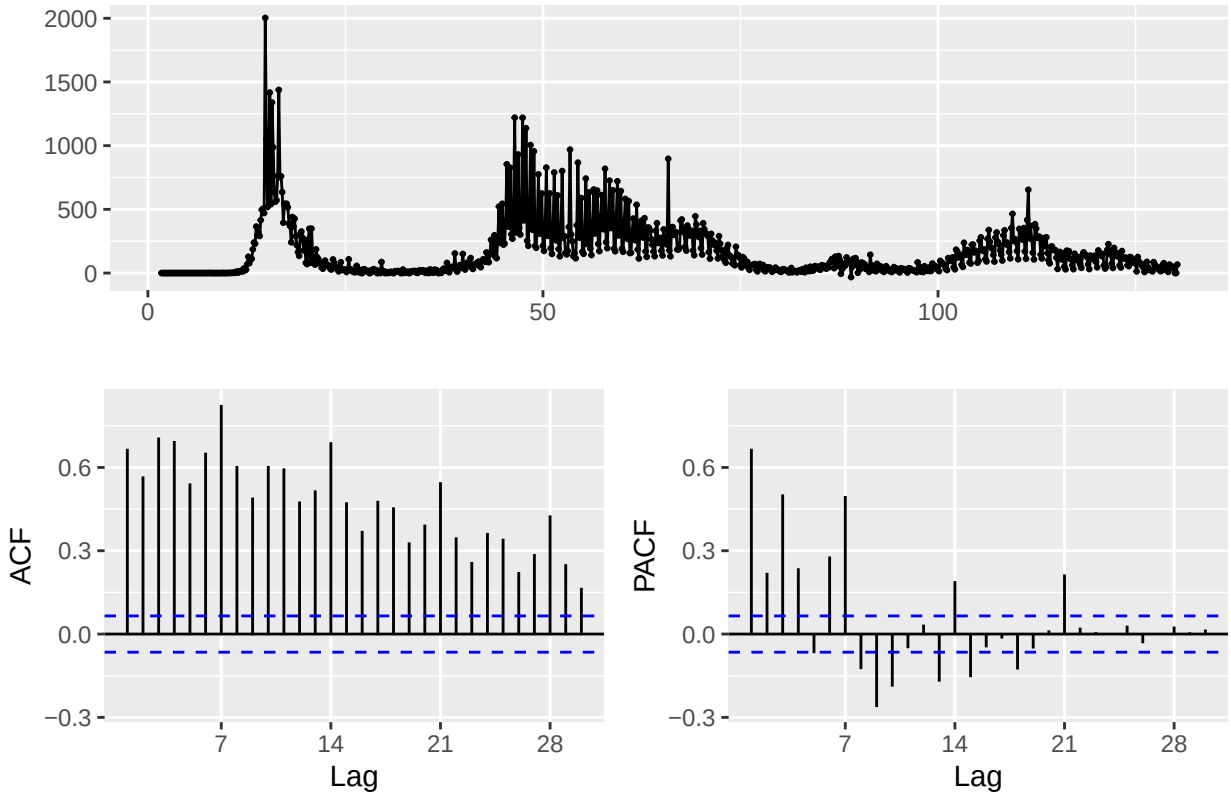
```
library(forecast)
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

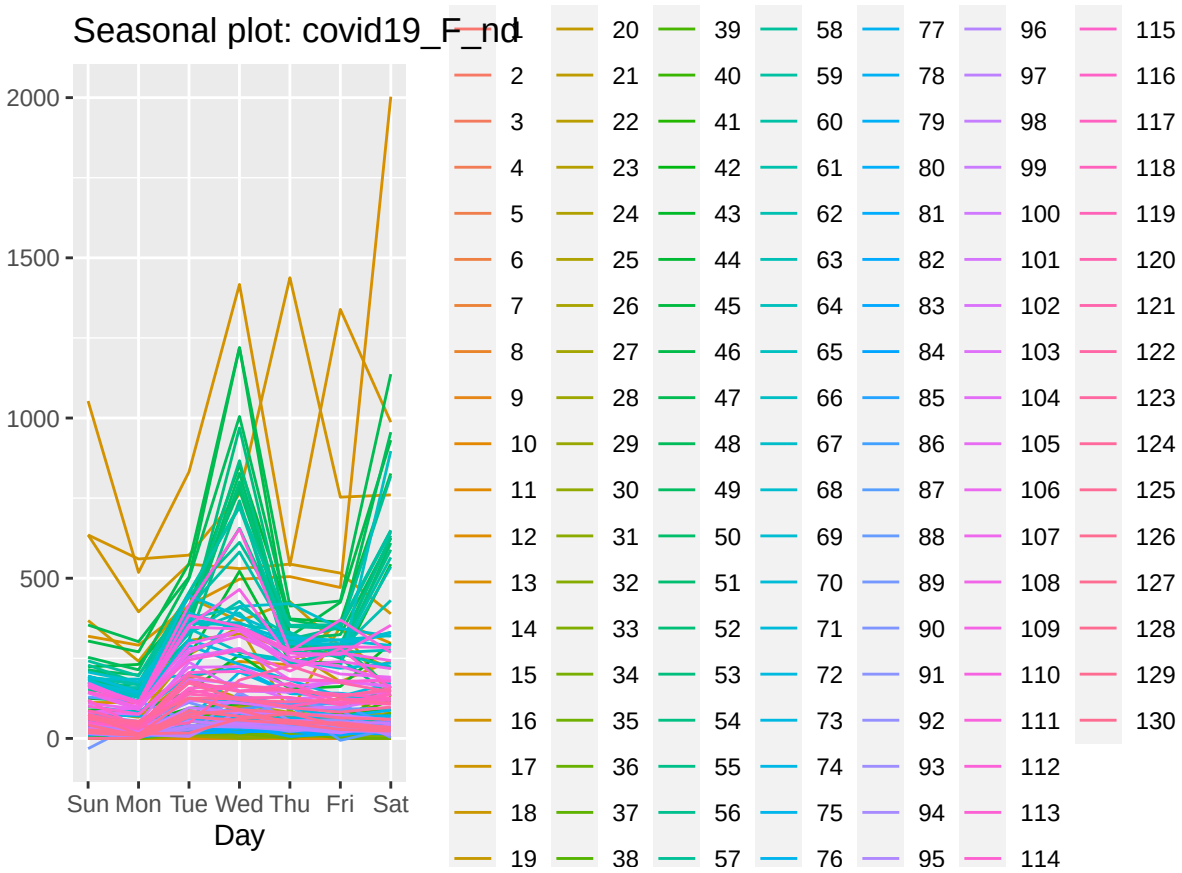
```
library(ggplot2)
autoplot(covid19_F_nd) +
ggtitle('Number of new deaths of Covid19 in France')+ xlab('days')+
ylab('Number of new deaths of Covid19')
```



```
ggtsdisplay(covid19_F_nd)
```

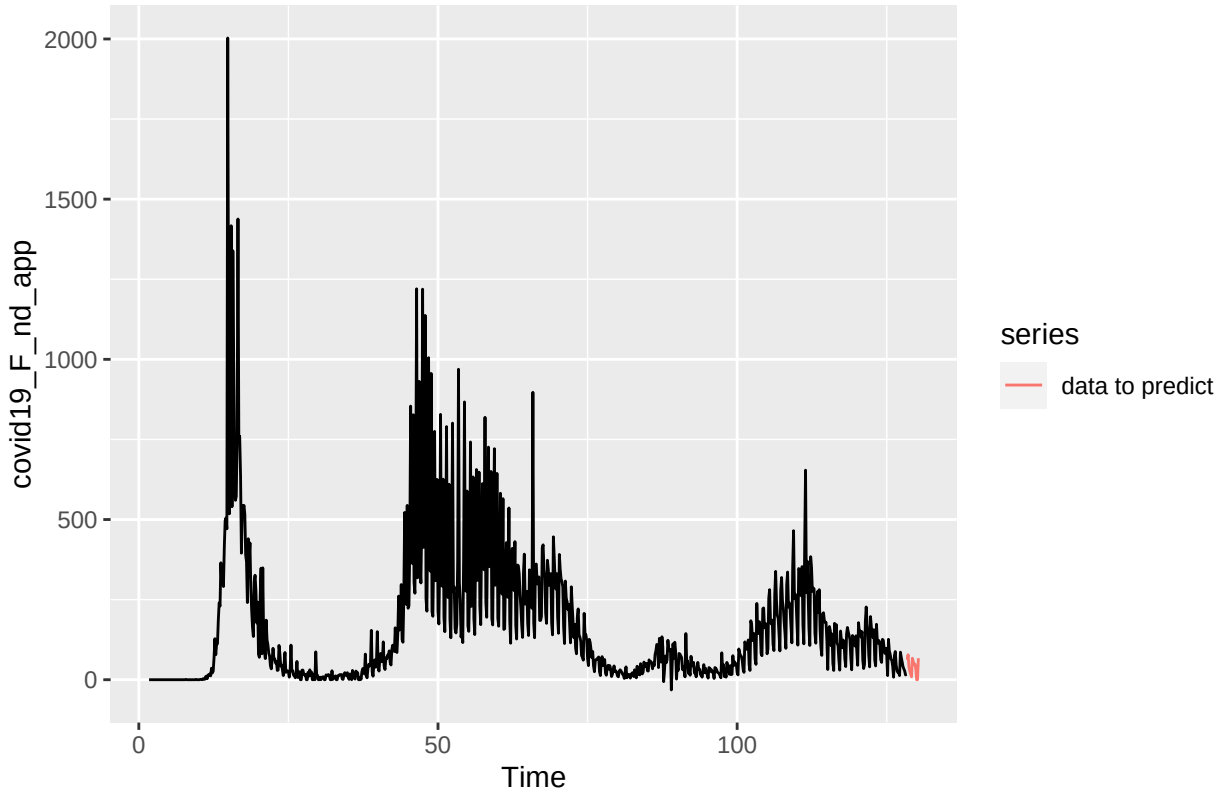



```
ggseasonplot(covid19_F_nd)
```



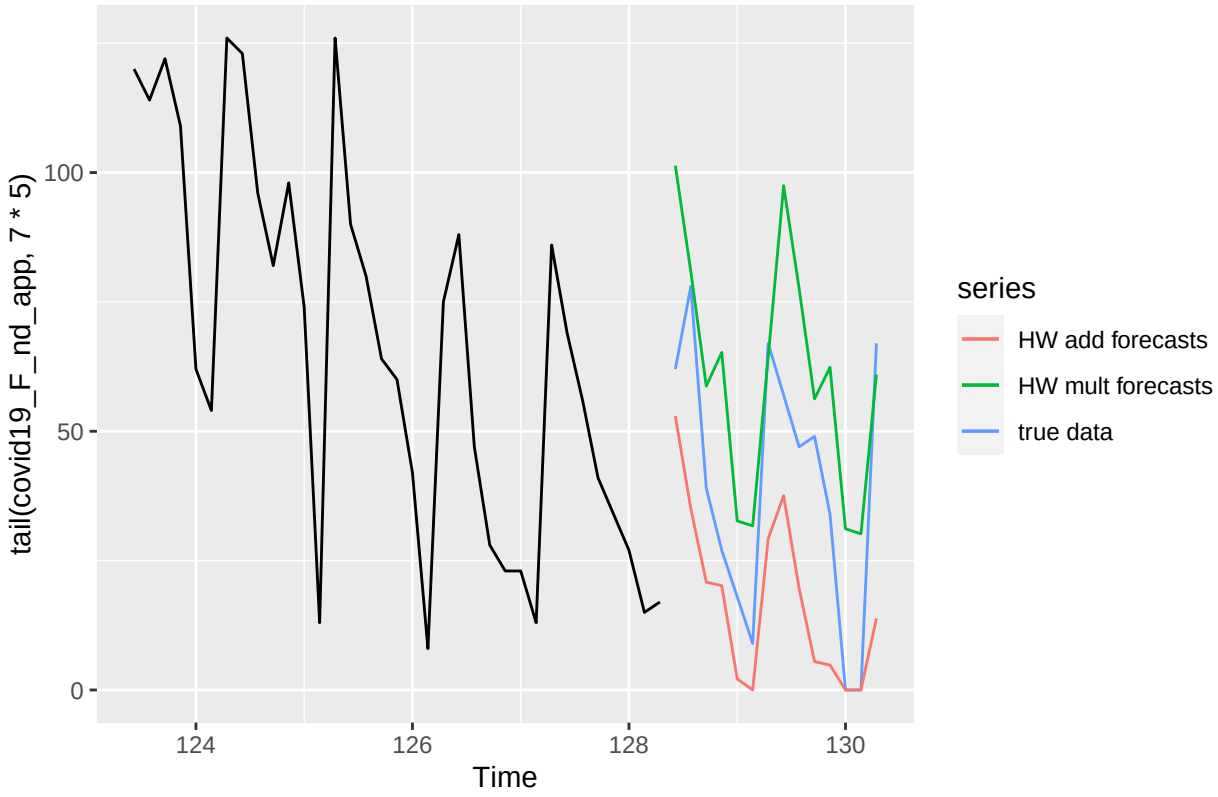
It looks like if there is a trend and a seasonnal pattern of period equal to 7 (week)

```
covid19_F_nd_app=head(covid19_F_nd,length(covid19_F_nd)-14)
covid19_F_nd_test=tail(covid19_F_nd,14)
autoplot(covid19_F_nd_app)+autolayer(covid19_F_nd_test, series="data to predict")
```



Forecasting of the number of new death with HW

```
fit=hw(covid19_F_nd_app)
tmp=covid19_F_nd_app;tmp[tmp<1]=1
fit2=hw(tmp,seasonal = "multiplicative")
prev=forecast(fit,h=14)
prev$mean[prev$mean<0]=0
prev2=forecast(fit2,h=14)
autoplot(tail(covid19_F_nd_app,7*5))+
autolayer(covid19_F_nd_test, series="true data")+
  autolayer(prev$mean, series="HW add forecasts")+
  autolayer(prev2$mean, series="HW mult forecasts")
```



Computing the MAPE

```
cat('RMSE for HW additive:',sqrt(mean((prev$mean-covid19_F_nd_test)^2)),'\n')
```

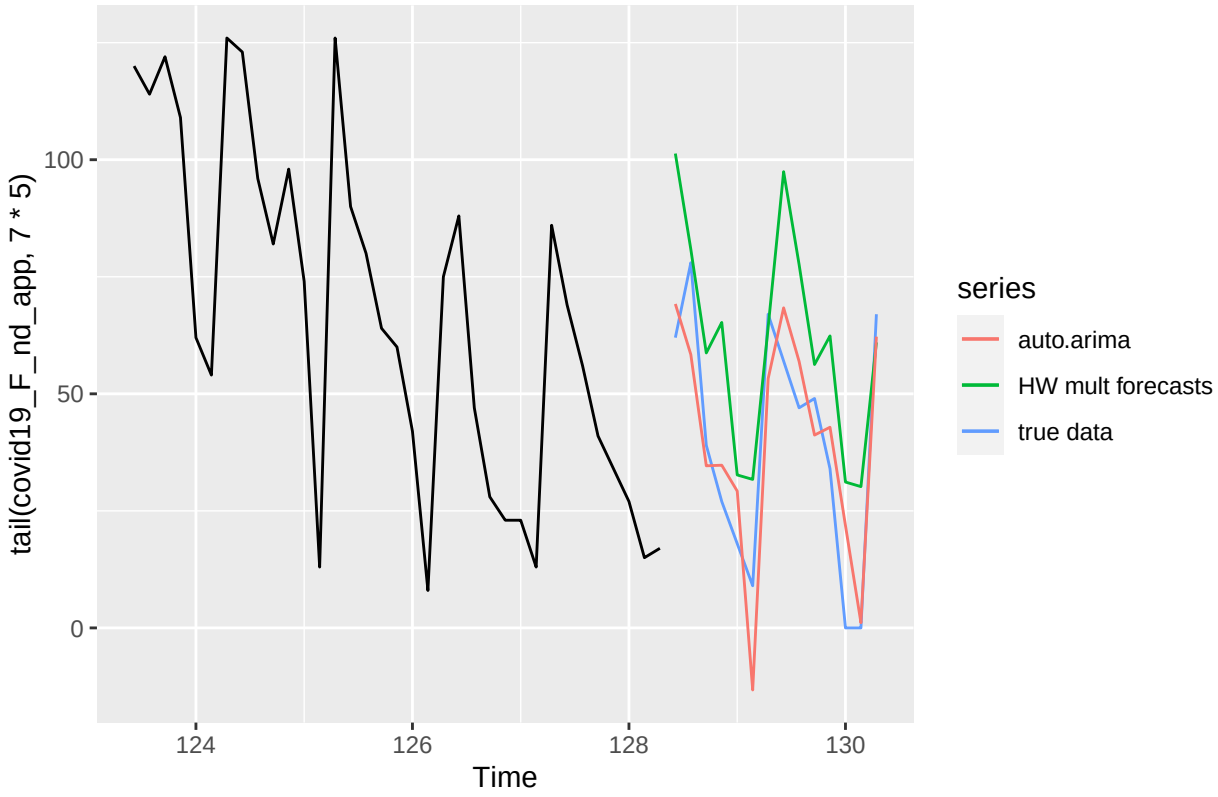
```
## RMSE for HW additive: 27.72868
```

```
cat('RMSE for HW multiplicative:',sqrt(mean((prev2$mean-covid19_F_nd_test)^2)),'\n')
```

```
## RMSE for HW multiplicative: 26.04138
```

Forecasting of the number of new cases

```
fit3=auto.arima(covid19_F_nd_app)
prev3=forecast(fit3,h=14)
autoplot(tail(covid19_F_nd_app,7*5))+
autolayer(covid19_F_nd_test, series="true data")+
  autolayer(prev2$mean, series="HW mult forecasts")+
  autolayer(prev3$mean, series="auto.arima")
```

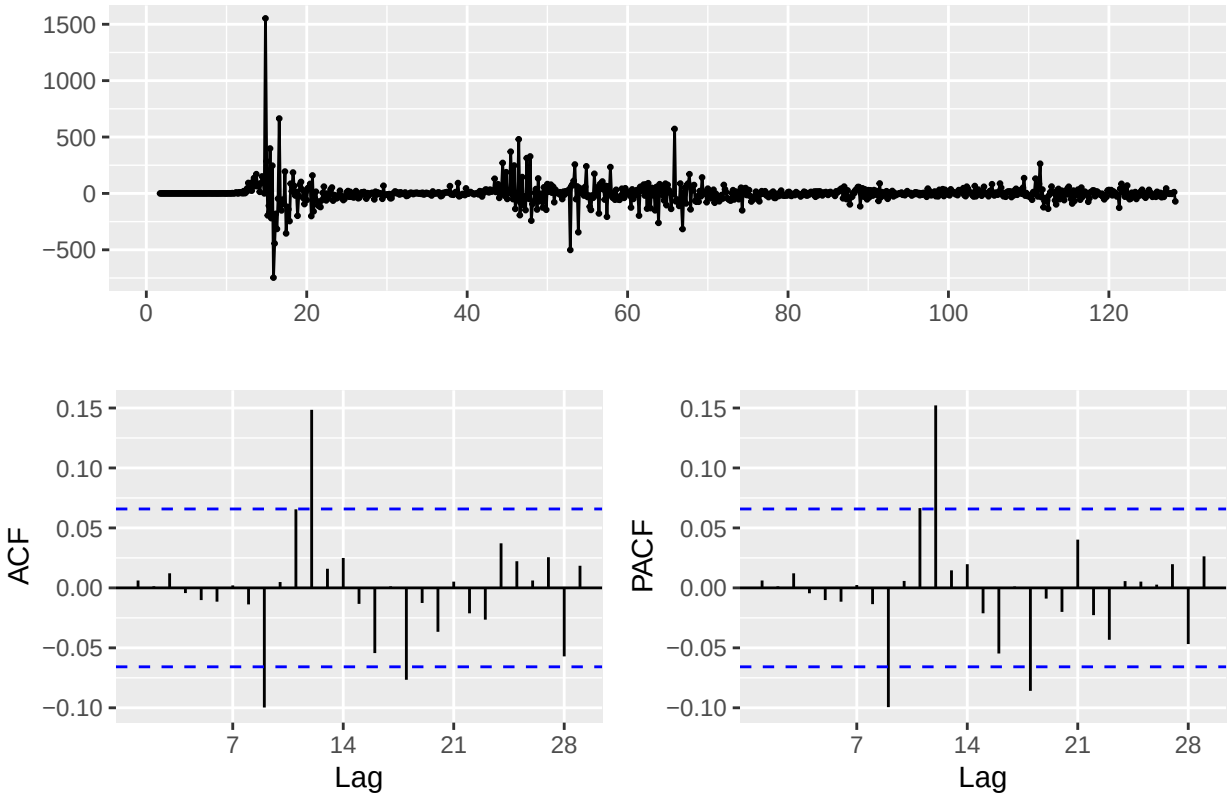



```
cat('RMSE for auto.arima:',sqrt(mean((prev3$mean-covid19_F_nd_test)^2)),'\n')
```

```
## RMSE for auto.arima: 12.52447
```

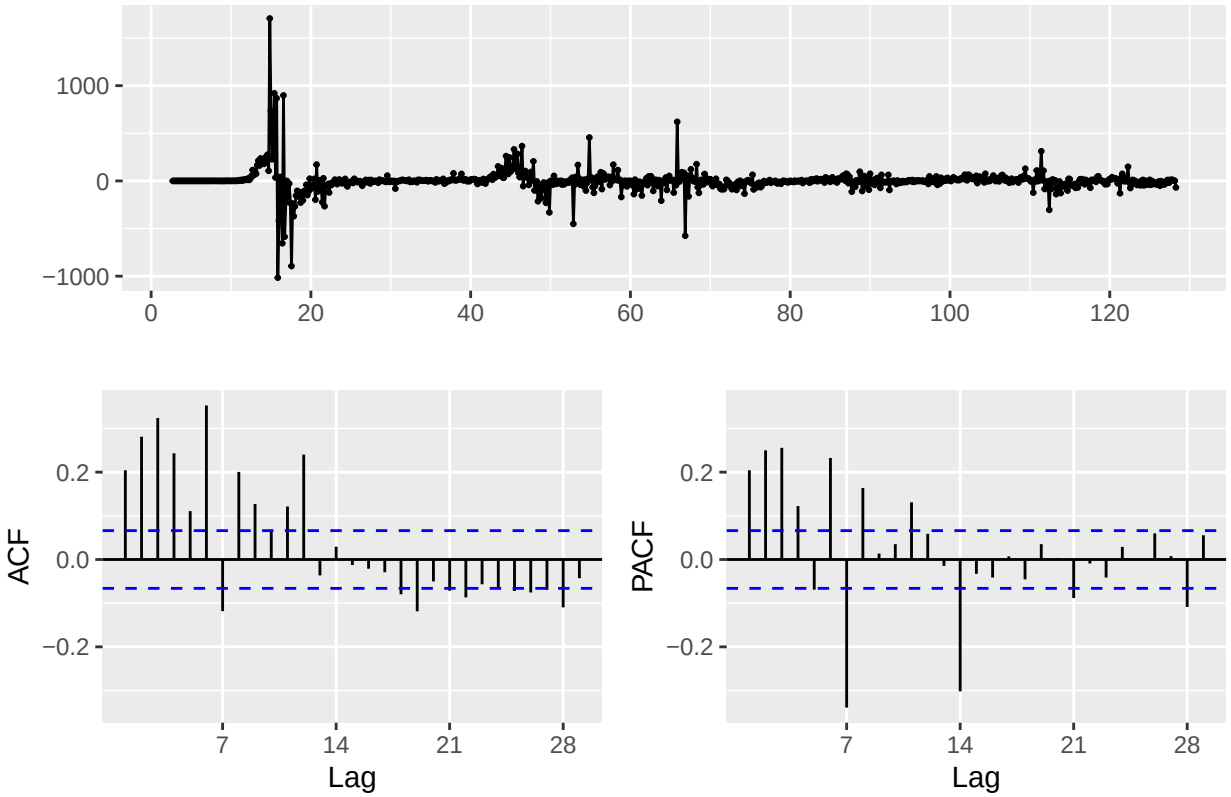
Let's try to improve the ARIMA modeling. Let's start by looking at the residual of the auto.arima modeling

```
ggtsdisplay(fit3$residuals)
```



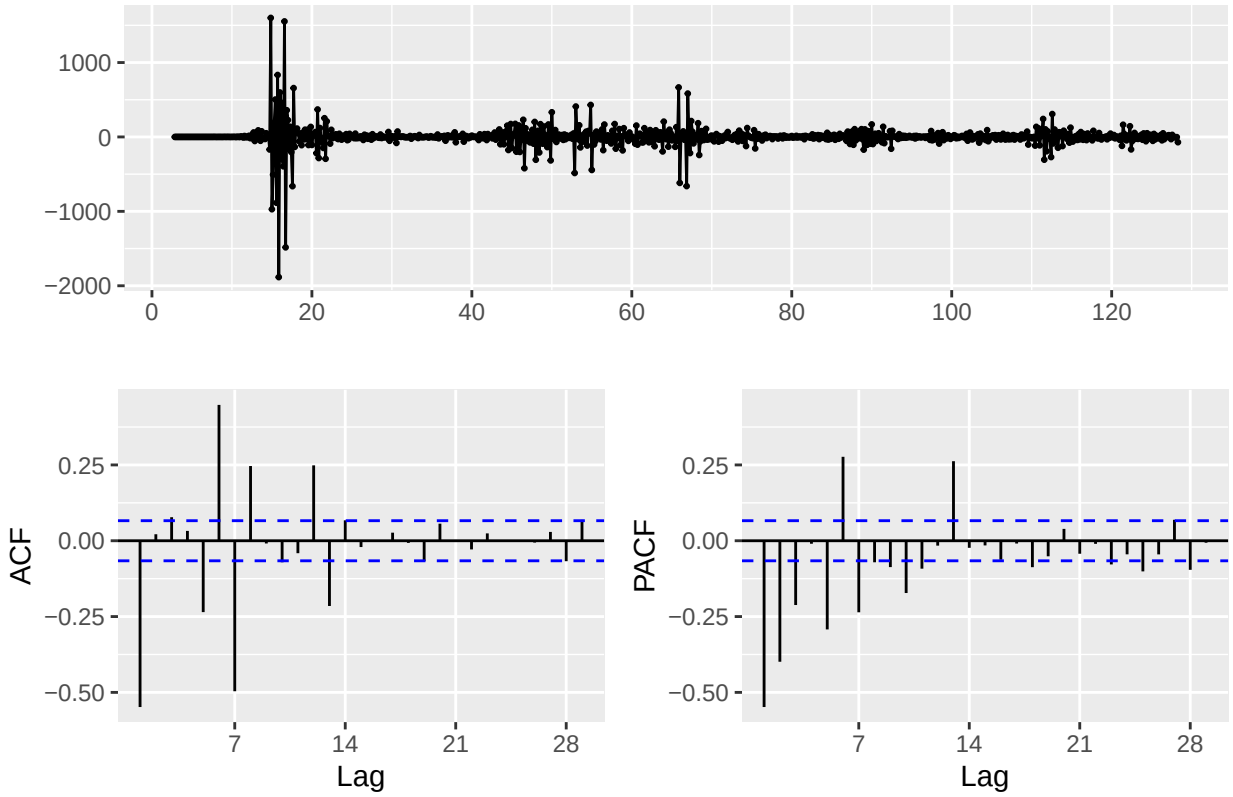
Let's differenciate the time serie

```
ggtsdisplay(diff(covid19_F_nd_app,lag=7))
```

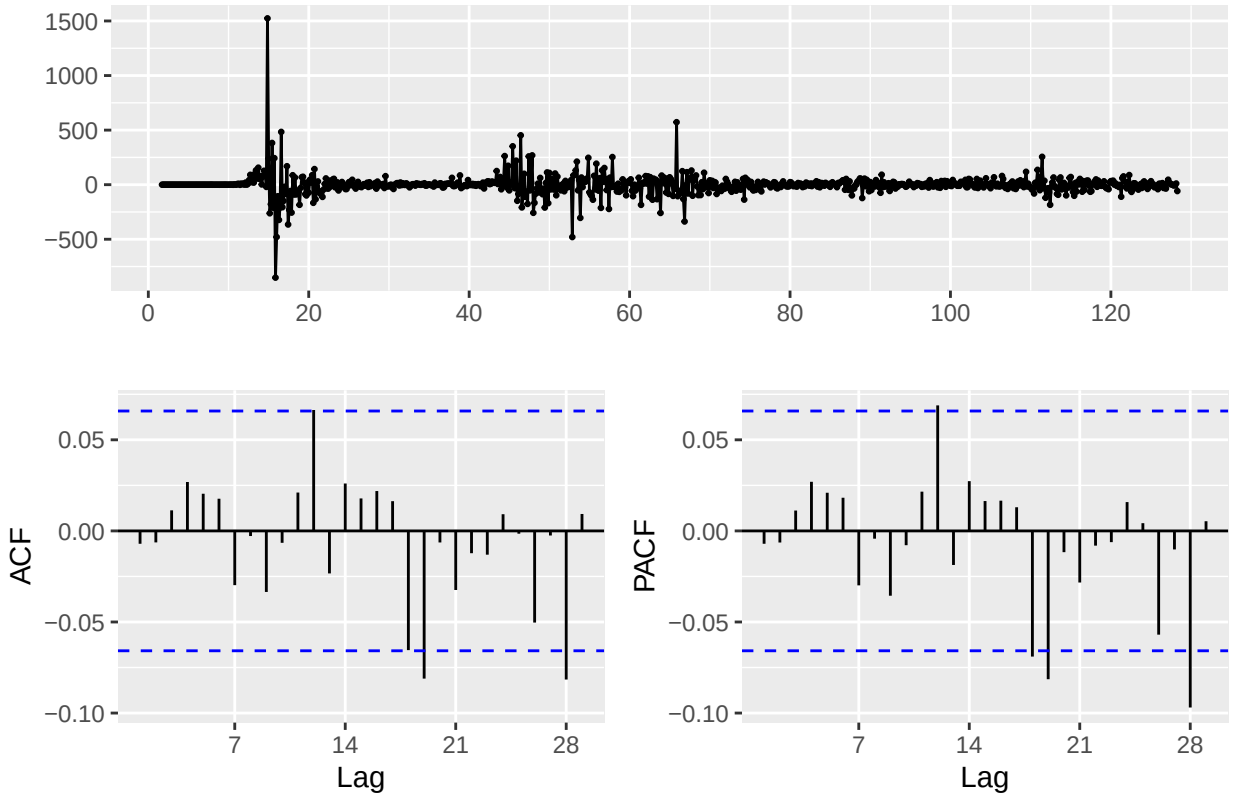


the ACF show that there is still a trend. We differenciate again

```
ggtsdisplay(diff(diff(covid19_F_nd_app,lag=7)))
```

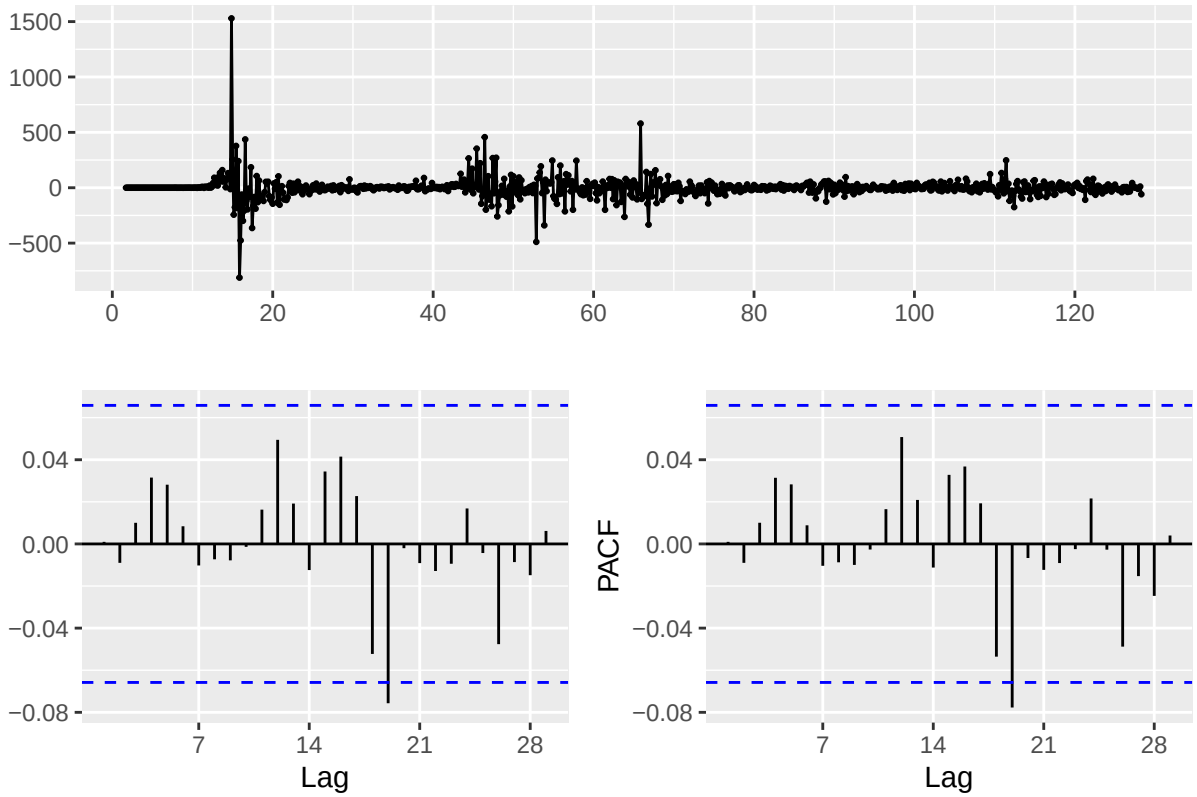



```
fit4=arima(covid19_F_nd_app,order =c(0,1,13) ,seasonal = c(0,1,0))
ggtsdisplay(fit4$residuals)
```

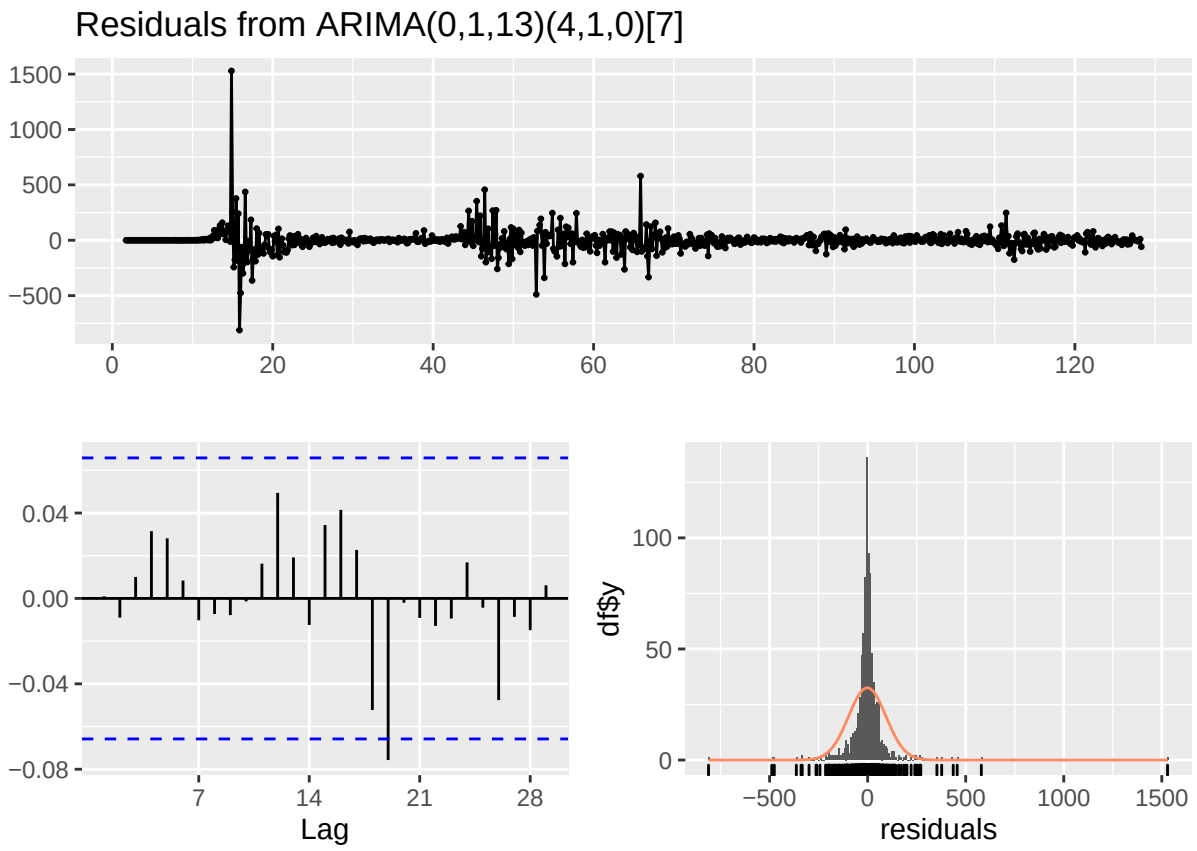


It remains significant ACF and PACF at lag 28. We try to add a seasonal AR of order 4

```
fit4=arima(covid19_F_nd_app,order =c(0,1,13) ,seasonal = c(4,1,0))
ggtsdisplay(fit4$residuals)
```



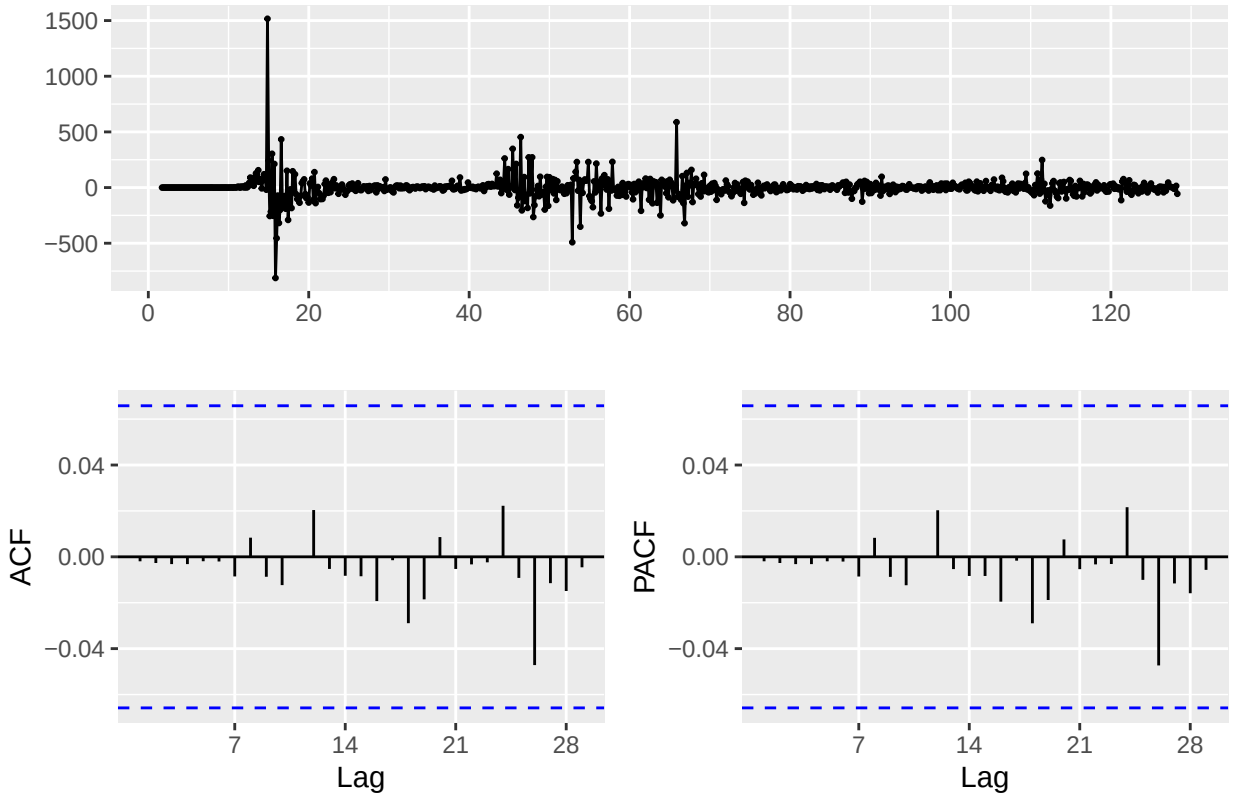

```
checkresiduals(fit4)
```



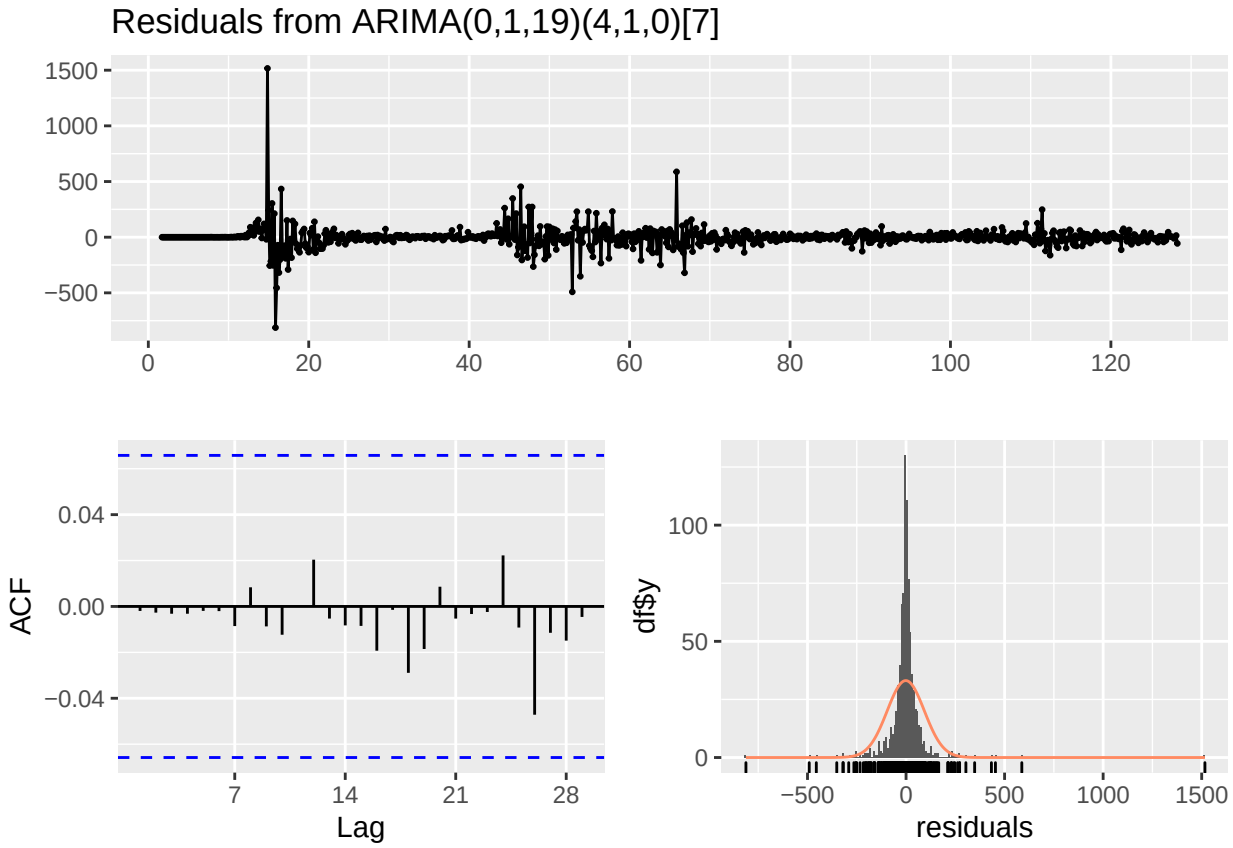
```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,1,13)(4,1,0)[7]
## Q* = 15.712, df = 3, p-value = 0.001299
##
## Model df: 17.   Total lags used: 20
```

It remains significant ACF and PACF at lag 19. We try to increase the MA order from 13 to 19

```
fit4=arima(covid19_F_nd_app,order =c(0,1,19) ,seasonal = c(4,1,0))
ggtsdisplay(fit4$residuals)
```



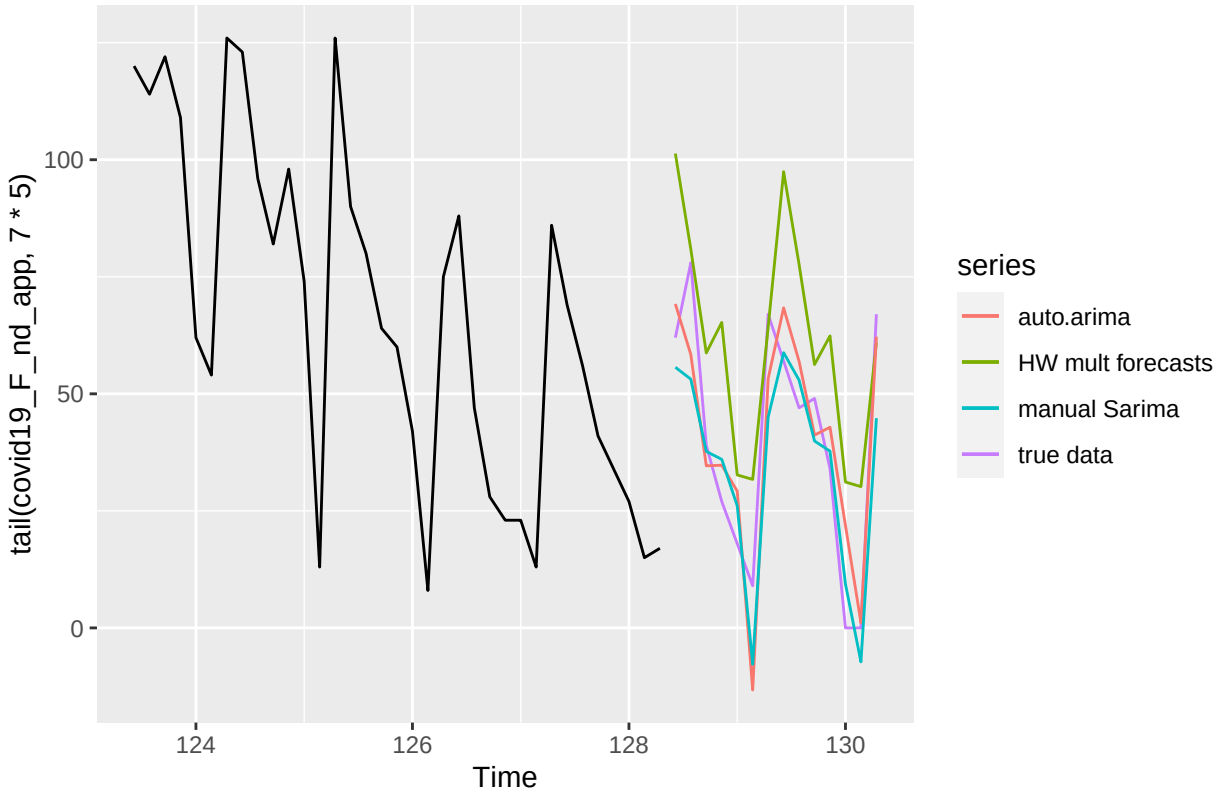

```
checkresiduals(fit4)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,1,19)(4,1,0)[7]
## Q* = 4.9756, df = 3, p-value = 0.1736
##
## Model df: 23.   Total lags used: 26
```

Let's try it for the forecast

```
prev4=forecast(fit4,h=14)
autoplot(tail(covid19_F_nd_app,7*5))+
autolayer(covid19_F_nd_test, series="true data")+
  autolayer(prev2$mean, series="HW mult forecasts")+
  autolayer(prev3$mean, series="auto.arima")+
  autolayer(prev4$mean, series="manual Sarima")
```




```
cat('RMSE for manual Sarima:',sqrt(mean((prev4$mean-covid19_F_nd_test)^2)),'\n')
```

```
## RMSE for manual Sarima: 12.92529
```

We can check the AIC for the 2 arima model

```
fit3$aic
```

```
## [1] 10576.91
```

```
fit4$aic
```

```
## [1] 10552.11
```