US Covid-19 Cases Time Series Analysis

Stefan Maciolek

5/28/2020

```r
#importing Covid Data from Johns Hopkins
uscases <-
read.csv(url("https://raw.githubusercontent.com/CSSEGISandData/COVID-
19/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_co
nfirmed_US.csv"))
usdeaths <-
read.csv(url("https://raw.githubusercontent.com/CSSEGISandData/COVID-
19/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_de
aths_US.csv"))
uscases <- uscases[,-c(1:5,8:11)]
usdeaths <- usdeaths[,-c(1:5,8:12)]
n <- ncol(uscases)-2
date <- 1:n
date <- as.Date(date,origin = "2020-01-21")
format(date,format = "%b %d %y")
```

```
##   [1] "Jan 22 20" "Jan 23 20" "Jan 24 20" "Jan 25 20" "Jan 26 20" "Jan 27
20"
##   [7] "Jan 28 20" "Jan 29 20" "Jan 30 20" "Jan 31 20" "Feb 01 20" "Feb 02
20"
##  [13] "Feb 03 20" "Feb 04 20" "Feb 05 20" "Feb 06 20" "Feb 07 20" "Feb 08
20"
##  [19] "Feb 09 20" "Feb 10 20" "Feb 11 20" "Feb 12 20" "Feb 13 20" "Feb 14
20"
##  [25] "Feb 15 20" "Feb 16 20" "Feb 17 20" "Feb 18 20" "Feb 19 20" "Feb 20
20"
##  [31] "Feb 21 20" "Feb 22 20" "Feb 23 20" "Feb 24 20" "Feb 25 20" "Feb 26
20"
##  [37] "Feb 27 20" "Feb 28 20" "Feb 29 20" "Mar 01 20" "Mar 02 20" "Mar 03
20"
##  [43] "Mar 04 20" "Mar 05 20" "Mar 06 20" "Mar 07 20" "Mar 08 20" "Mar 09
20"
##  [49] "Mar 10 20" "Mar 11 20" "Mar 12 20" "Mar 13 20" "Mar 14 20" "Mar 15
20"
##  [55] "Mar 16 20" "Mar 17 20" "Mar 18 20" "Mar 19 20" "Mar 20 20" "Mar 21
20"
##  [61] "Mar 22 20" "Mar 23 20" "Mar 24 20" "Mar 25 20" "Mar 26 20" "Mar 27
20"
##  [67] "Mar 28 20" "Mar 29 20" "Mar 30 20" "Mar 31 20" "Apr 01 20" "Apr 02
20"
##  [73] "Apr 03 20" "Apr 04 20" "Apr 05 20" "Apr 06 20" "Apr 07 20" "Apr 08
20"
```

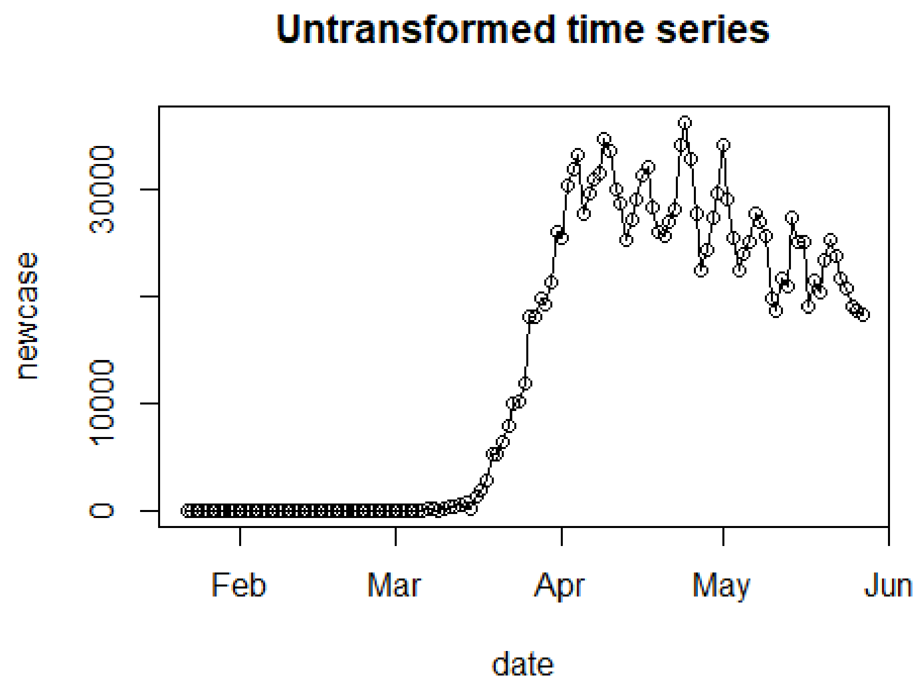
```
##  [79] "Apr 09 20" "Apr 10 20" "Apr 11 20" "Apr 12 20" "Apr 13 20" "Apr 14
20"
##  [85] "Apr 15 20" "Apr 16 20" "Apr 17 20" "Apr 18 20" "Apr 19 20" "Apr 20
20"
##  [91] "Apr 21 20" "Apr 22 20" "Apr 23 20" "Apr 24 20" "Apr 25 20" "Apr 26
20"
##  [97] "Apr 27 20" "Apr 28 20" "Apr 29 20" "Apr 30 20" "May 01 20" "May 02
20"
## [103] "May 03 20" "May 04 20" "May 05 20" "May 06 20" "May 07 20" "May 08
20"
## [109] "May 09 20" "May 10 20" "May 11 20" "May 12 20" "May 13 20" "May 14
20"
## [115] "May 15 20" "May 16 20" "May 17 20" "May 18 20" "May 19 20" "May 20
20"
## [121] "May 21 20" "May 22 20" "May 23 20" "May 24 20" "May 25 20" "May 26
20"
## [127] "May 27 20"
```

The data were imported from Johns Hopkins, cleaned up, and a date variable was created for plotting purposes. For this analysis the death data wasn't used, just the case numbers

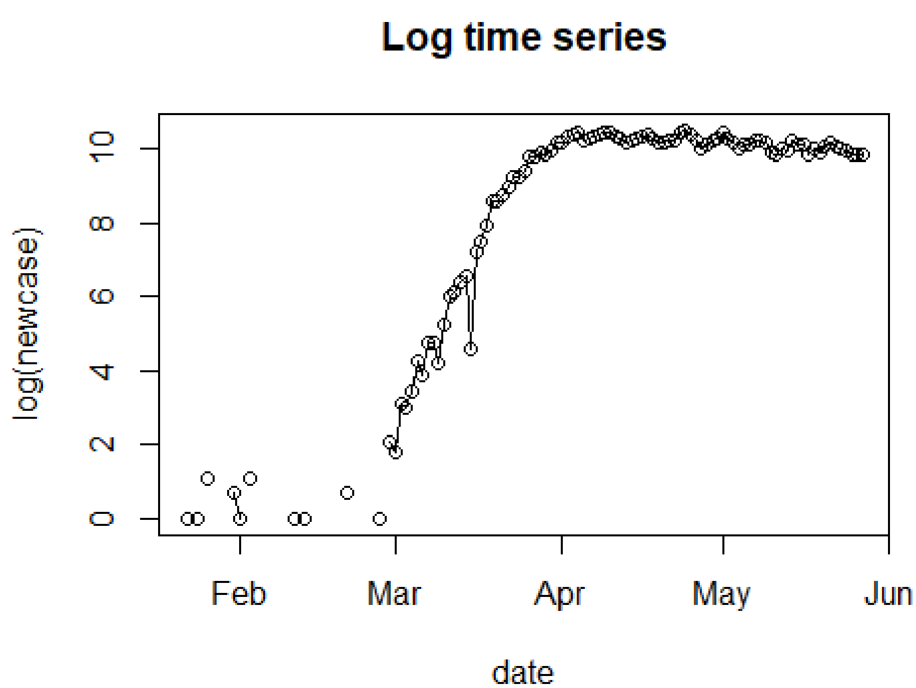
```
#finding total cases and turning them into a time series object
library(TSA)

##
## Attaching package: 'TSA'

## The following objects are masked from 'package:stats':
##
##     acf, arima

## The following object is masked from 'package:utils':
##
##     tar

totcase <- colSums(uscases[,3:(n+2)])
newcase <- rep(0,n)
newcase[1] <- totcase[1]
newcase[2:n] <- diff(totcase)
newcase <- ts(data=newcase,start=c(2020,01,22),frequency = 365)
```

A the data were turned into a time series object

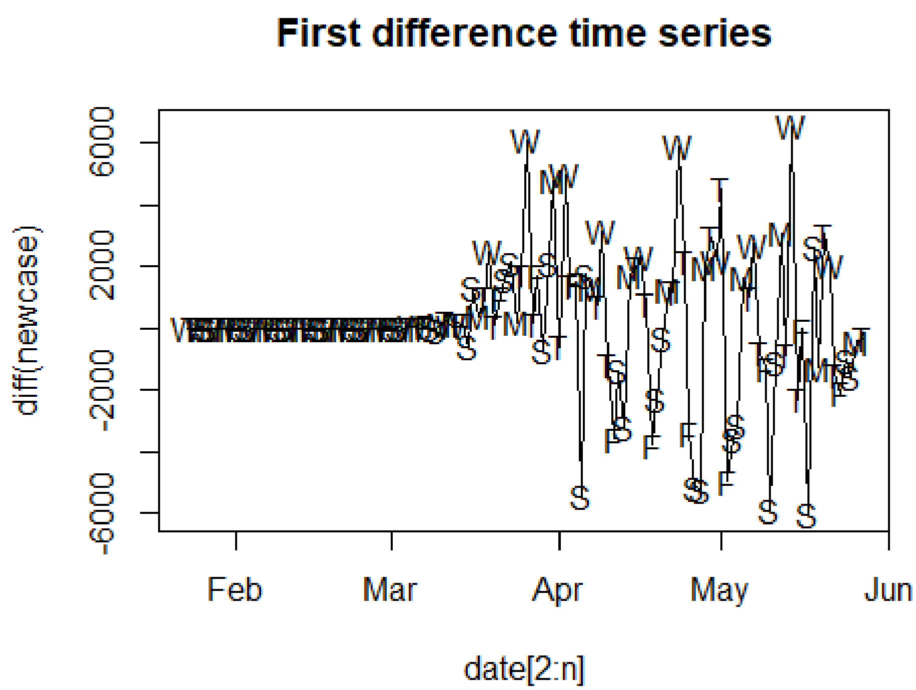
```
#trying various transformations
wkday <- c("W","T","F","S","S","M","T")
plot(date,newcase,type="o",main="Untransformed time series")
```



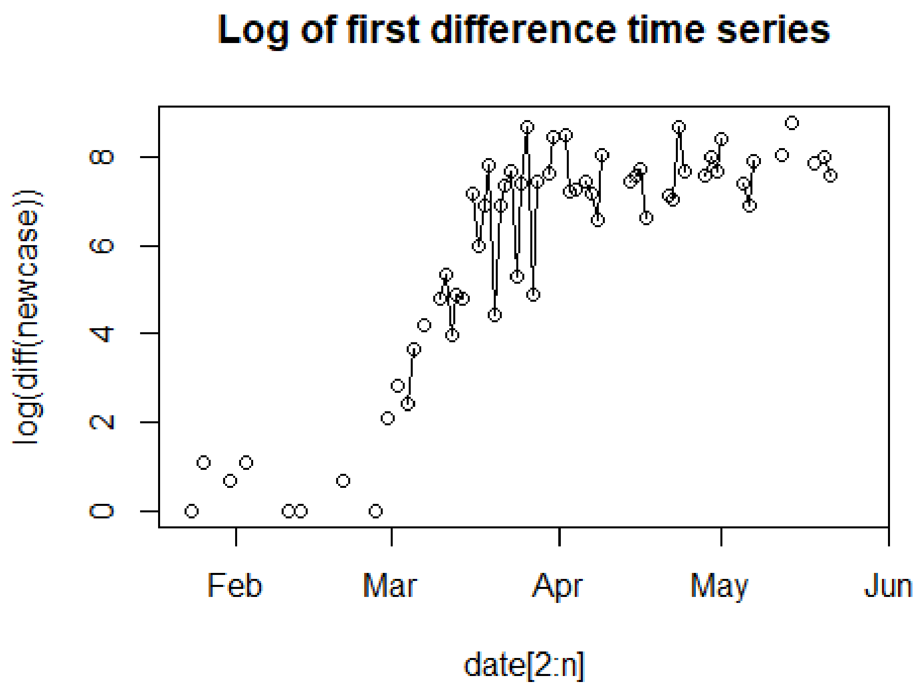

```
plot(date,log(newcase),type="o",main="Log time series")
```



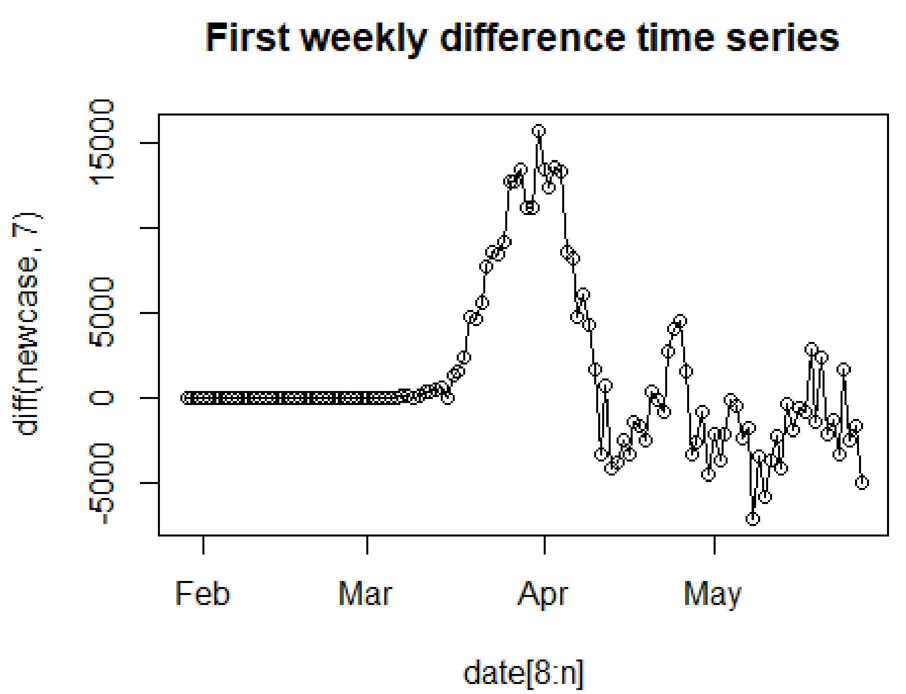
```r
plot(date[2:n],diff(newcase),type="o",main="First difference time series",pch=wkday)
```



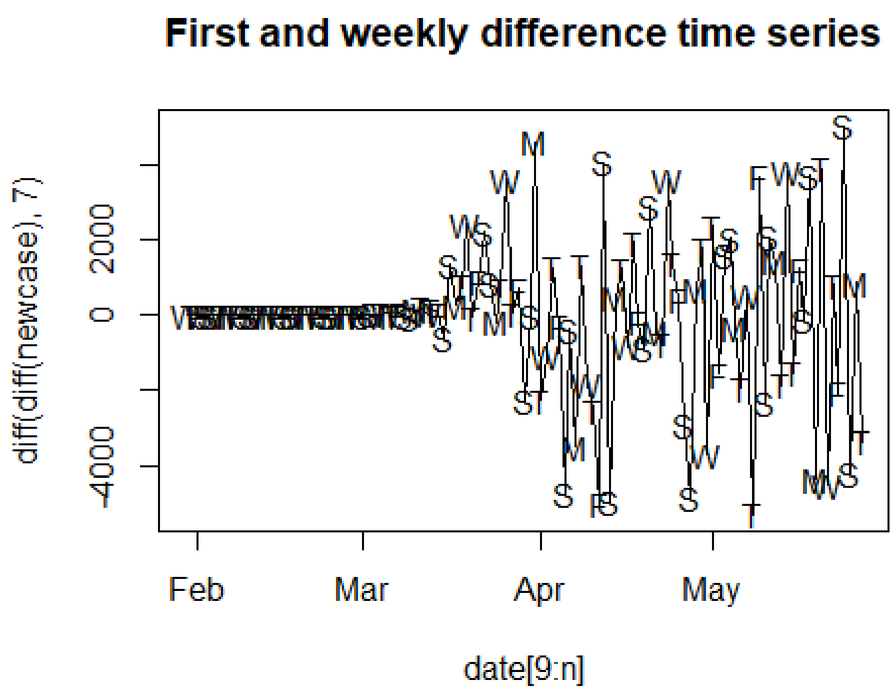

```r
plot(date[2:n],log(diff(newcase)),type="o",main="Log of first difference time series")

## Warning in log(diff(newcase)): NaNs produced
```




```r
plot(date[8:n],diff(newcase,7),type="o",main="First weekly difference time
series")
```

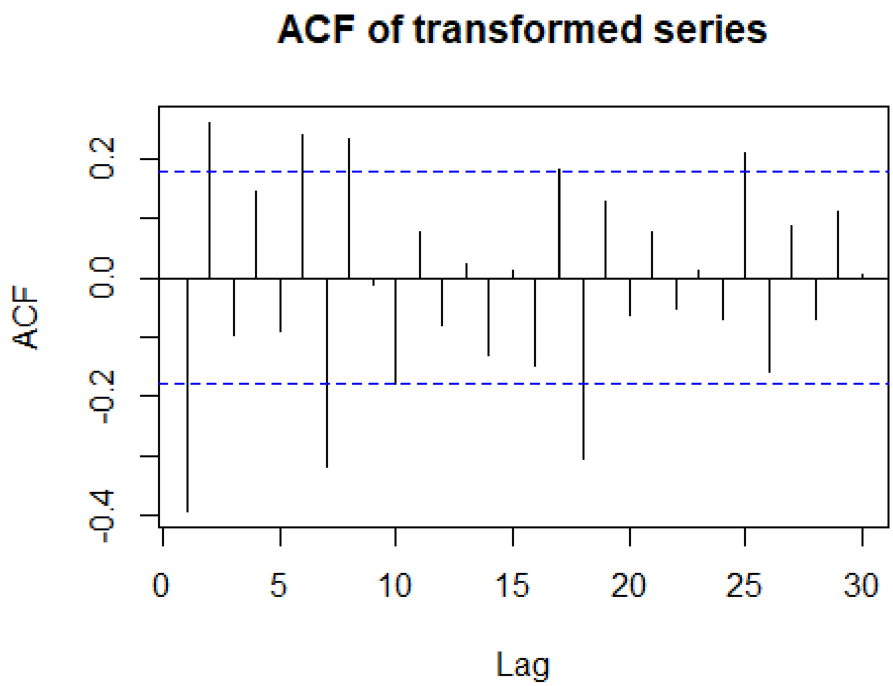


```r
plot(date[9:n],diff(diff(newcase),7),type="o",main="First and weekly
difference time series",pch=wkday)
```

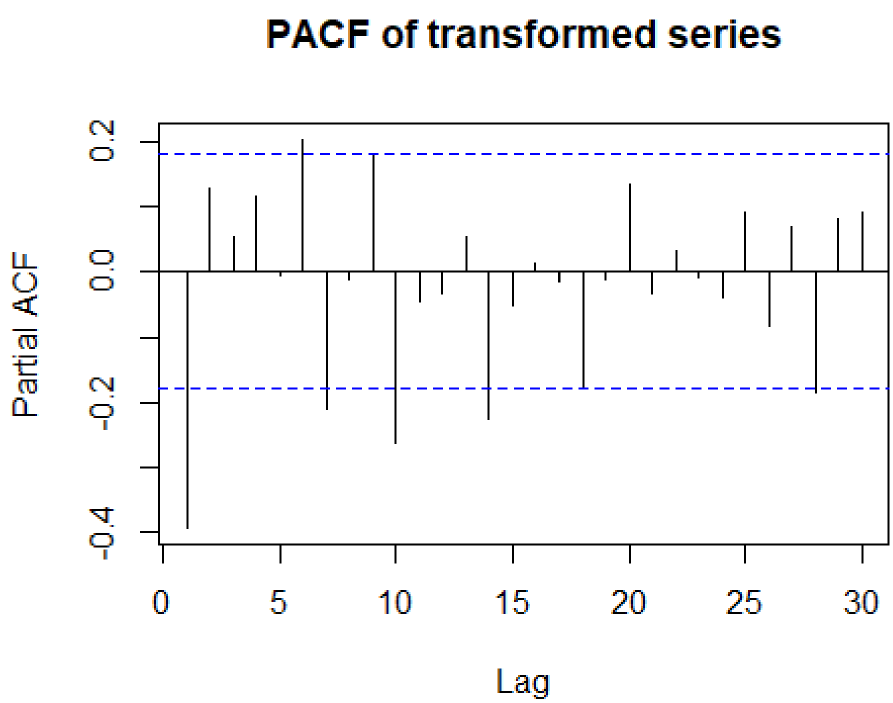


Several transformations and combinations of transformations were tested to see if they made the data random. In the end, a first and first seasonal difference with a weekly period were chosen.

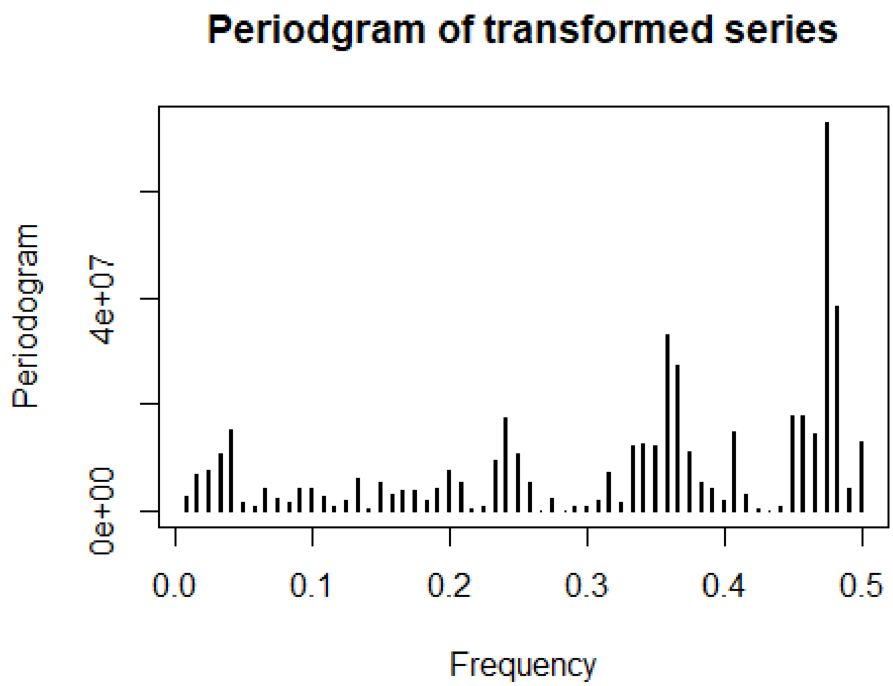
```r
#starting with an ARIMA model with both a first and first weekly difference
#preliminary analysis
transcase <- diff(diff(newcase),7)
acf(as.vector(transcase),main="ACF of transformed series",lag.max=30)
```




```
pacf(as.vector(transcase),main="PACF of transformed series",lag.max=30)
```




```
periodogram(transcase,main="Periodgram of transformed series")
```



The ACF and PACF of the series suggest a starting model with at least 1 AR and 1 seasonal AR component.

```
#model fitting:
#looking at the periodograms, it appears a ARIMA(1,1,0)x(1,1,0)_7 is
justified to start
tsmodel <- function(ar,ma,ars,mas){
  mod <-
arima(newcase,order=c(ar,1,ma),seasonal=list(order=c(ars,1,mas),period=7))
  return(mod)
}
tsmodel(1,0,1,0)

##
## Call:
## arima(x = newcase, order = c(ar, 1, ma), seasonal = list(order = c(ars, 1,
mas),
##     period = 7))
##
## Coefficients:
##           ar1     sar1
##       -0.3344  -0.2568
## s.e.   0.0904   0.0953
##
## sigma^2 estimated as 3188580:  log likelihood = -1060.17,  aic = 2124.34

tsmodel(2,0,1,0)
```

```
##
## Call:
## arima(x = newcase, order = c(ar, 1, ma), seasonal = list(order = c(ars, 1,
mas),
##     period = 7))
##
## Coefficients:
##          ar1     ar2      sar1
##       -0.259  0.1927  -0.2972
## s.e.   0.094  0.0912   0.0934
##
## sigma^2 estimated as 3067794:  log likelihood = -1058,  aic = 2121.99
```

```
tsmodel(2,0,2,0)
```

```
##
## Call:
## arima(x = newcase, order = c(ar, 1, ma), seasonal = list(order = c(ars, 1,
mas),
##     period = 7))
##
## Coefficients:
##           ar1     ar2      sar1      sar2
##       -0.2424  0.1793  -0.3636  -0.2169
## s.e.   0.0968  0.0934   0.0991   0.0936
##
## sigma^2 estimated as 2920166:  log likelihood = -1055.39,  aic = 2118.78
```

```
tsmodel(3,0,2,0) #not significant
```

```
##
## Call:
## arima(x = newcase, order = c(ar, 1, ma), seasonal = list(order = c(ars, 1,
mas),
##     period = 7))
##
## Coefficients:
##           ar1     ar2     ar3      sar1      sar2
##       -0.2503  0.1866  0.0356  -0.3584  -0.2227
## s.e.   0.0988  0.0953  0.0948   0.1000   0.0950
##
## sigma^2 estimated as 2916556:  log likelihood = -1055.32,  aic = 2120.64
```

```
tsmodel(2,0,3,0) #not significant
```

```
##
## Call:
## arima(x = newcase, order = c(ar, 1, ma), seasonal = list(order = c(ars, 1,
mas),
##     period = 7))
##
```

```
## Coefficients:
##           ar1     ar2     sar1     sar2     sar3
##       -0.2408  0.1834  -0.3789  -0.2390  -0.0602
## s.e.   0.0968  0.0934   0.1021   0.1003   0.0980
##
## sigma^2 estimated as 2908774:  log likelihood = -1055.2,  aic = 2120.4
```

```
tsmodel(2,1,2,0)
```

```
##
## Call:
## arima(x = newcase, order = c(ar, 1, ma), seasonal = list(order = c(ars, 1,
mas),
##     period = 7))
##
## Coefficients:
##          ar1     ar2      ma1     sar1     sar2
##       0.4163  0.3484  -0.6987  -0.3566  -0.2261
## s.e.  0.1863  0.0882   0.1781   0.0976   0.0952
##
## sigma^2 estimated as 2892421:  log likelihood = -1054.78,  aic = 2119.55
```

```
tsmodel(2,1,2,1)
```

```
##
## Call:
## arima(x = newcase, order = c(ar, 1, ma), seasonal = list(order = c(ars, 1,
mas),
##     period = 7))
##
## Coefficients:
##           ar1      ar2     ma1    sar1    sar2     sma1
##       -0.8498  -0.0243  0.6236  0.5177  0.1189  -0.9300
## s.e.   0.2995   0.1561  0.2723  0.2988  0.1828   0.3759
##
## sigma^2 estimated as 2766731:  log likelihood = -1053.9,  aic = 2119.8
```

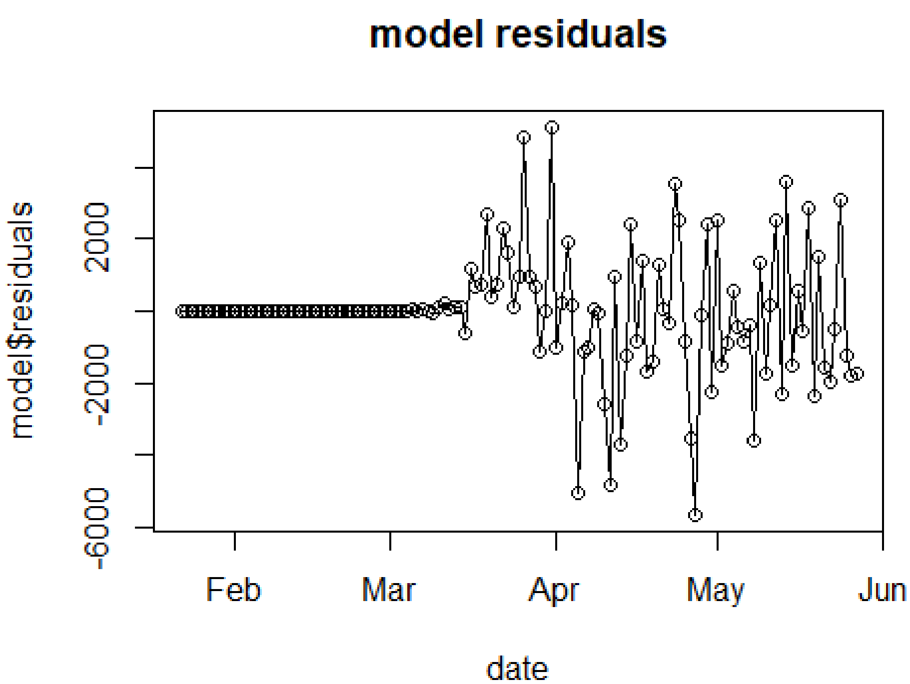
```
tsmodel(2,2,2,1) #not significant
```

```
##
## Call:
## arima(x = newcase, order = c(ar, 1, ma), seasonal = list(order = c(ars, 1,
mas),
##     period = 7))
##
## Coefficients:
##          ar1     ar2      ma1      ma2    sar1    sar2     sma1
##       0.1066  0.6321  -0.3125  -0.3817  0.4227  0.1037  -0.8495
## s.e.  0.2633  0.2231   0.3131   0.2779  0.3304  0.2050   0.3139
##
## sigma^2 estimated as 2791097:  log likelihood = -1053.48,  aic = 2120.95
```

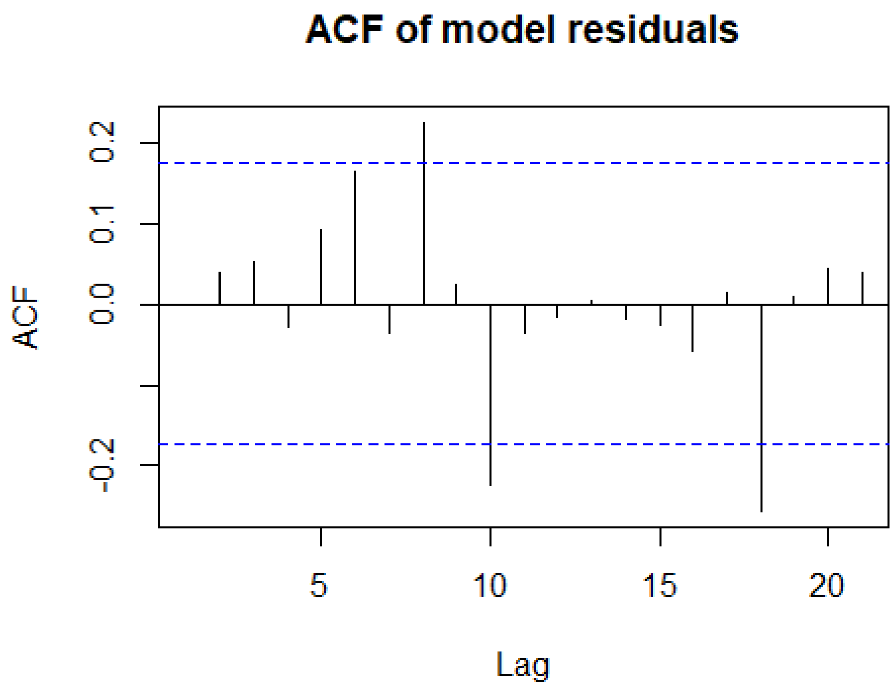
```r
tsmodel(2,1,2,2) #R did not like this

##
## Call:
## arima(x = newcase, order = c(ar, 1, ma), seasonal = list(order = c(ars, 1,
mas),
##     period = 7))
##
## Coefficients:

## Warning in sqrt(diag(x$var.coef)): NaNs produced

##           ar1      ar2     ma1     sar1   sar2     sma1    sma2
##       -0.8647  -0.0320  0.6393  -0.1223  0.251  -0.2674  -0.426
## s.e.   0.3931   0.1945  0.3479      NaN  0.185      NaN     NaN
##
## sigma^2 estimated as 2821964:  log likelihood = -1054.1,  aic = 2122.2

#the final model is an ARIMA(2,1,2)X(2,1,1)_7
```

The order of the model was raised one component at a time until adding further components were no longer significant. The final model was an ARIMA(2,1,1)X(2,1,2)_7

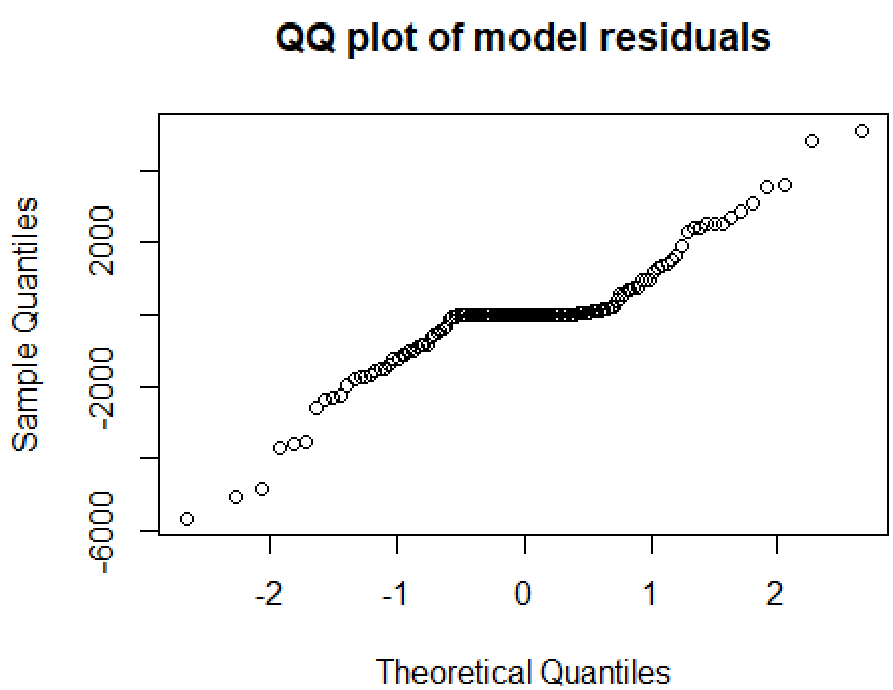
```r
#diagnostics:
model<-tsmodel(2,1,2,2)
plot(date,model$residuals,main="model residuals",type="o")
```



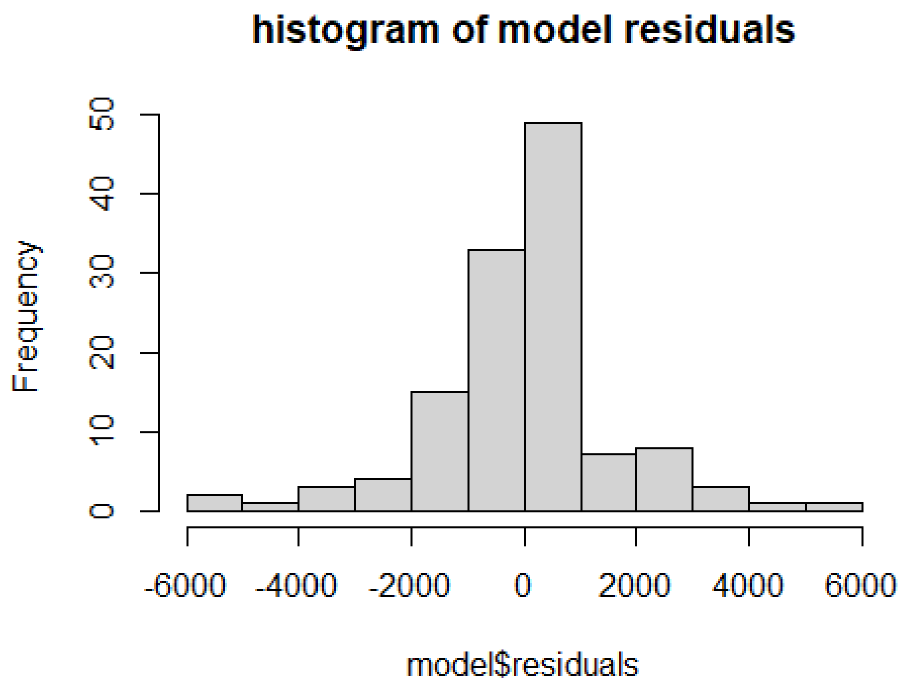

```r
acf(as.vector(model$residuals),main="ACF of model residuals")
```




```
qqnorm(model$residuals,main="QQ plot of model residuals")
```

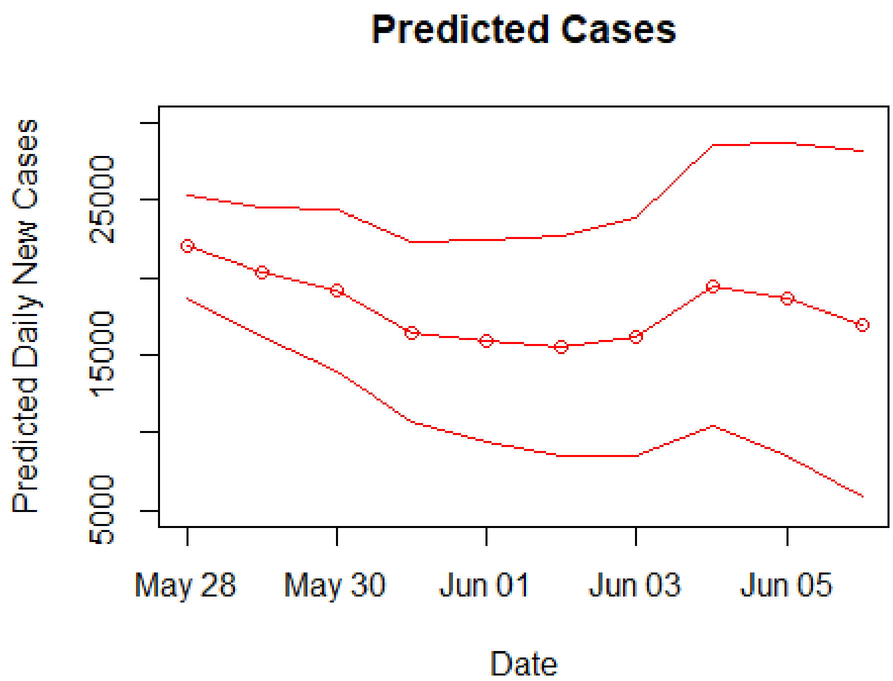



```
hist(model$residuals,main="histogram of model residuals")
```

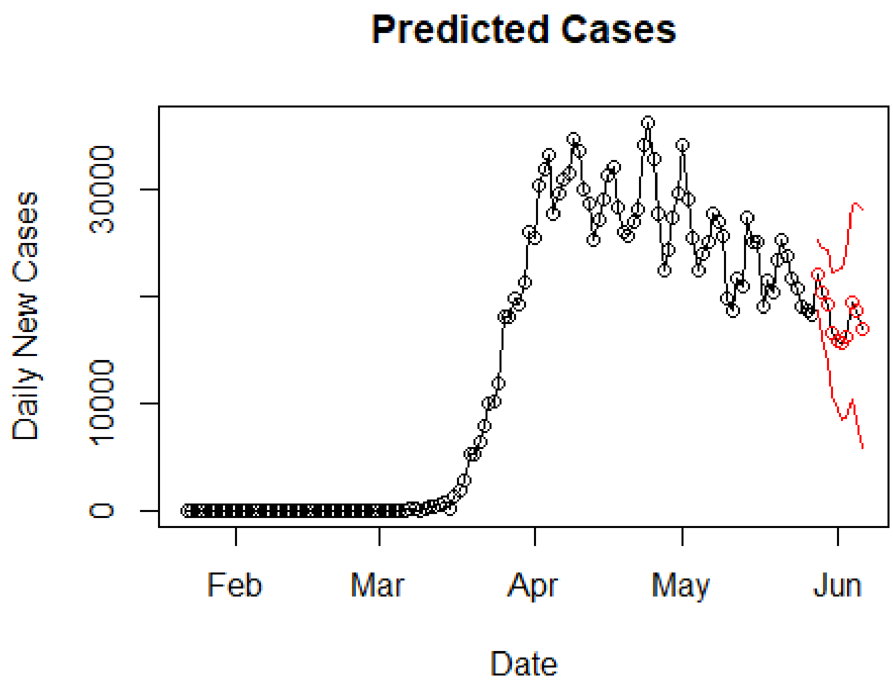


The residuals were analyzed and diagnostics produced. The residuals are more or less normal, and although they are somewhat autocorrelated at a few higher lags, it doesn't appear large enough to justify a more complex model

```
#predictions
predictions <- predict(model,n.ahead=10)
dateahead<- seq(from=(n+1),to=(n+10))
dateahead <- as.Date(dateahead,origin = "2020-01-21")
format(dateahead,format = "%b %d %y")

##  [1] "May 28 20" "May 29 20" "May 30 20" "May 31 20" "Jun 01 20" "Jun 02
20"
##  [7] "Jun 03 20" "Jun 04 20" "Jun 05 20" "Jun 06 20"

plot(dateahead,predictions$pred,main="Predicted
Cases",type="o",ylim=c(5000,30000),ylab="Predicted Daily New
Cases",xlab="Date",col="red")
lines(dateahead,(predictions$pred - 1.96*predictions$se),type="l",col="red")
lines(dateahead,(predictions$pred + 1.96*predictions$se),type="l",col="red")
```




```
datenew <- c(date,dateahead)
total <- c(newcase,predictions$pred)
plot(datenew,total,col=c(rep("black",n),rep("red",10)),main="Predicted
Cases",type="o",ylab="Daily New Cases",xlab="Date")
lines(dateahead,(predictions$pred - 1.96*predictions$se),type="l",col="red")
lines(dateahead,(predictions$pred + 1.96*predictions$se),type="l",col="red")
```



The model was used to create predictions for 10 days out.