Gonorrhoea Mobility ARIMA

Sulyok

3/8/2021

```
spearmanf<-function(x) {
  cor.test(gon_and_activity$goncases, x)$estimate
}
library(readr)
gon_and_activity <- read_csv("~/Downloads/gon_master.xlsx -
Sheet2.csv", col_types = cols(date = col_date(format = "%d/%m/%Y")))

summary(gon_and_activity)

##       date               goncases     drivingweekmean
transitweekmean
##  Min.   :2020-02-16   Min.   : 4.00   Min.   : 46.14   Min.   :
31.18
##  1st Qu.:2020-05-03   1st Qu.: 8.00   1st Qu.: 84.34   1st Qu.:
77.82
##  Median :2020-07-19   Median :10.00   Median :111.54   Median
:106.48
##  Mean   :2020-07-19   Mean   :11.02   Mean   :109.02   Mean   :
98.72
##  3rd Qu.:2020-10-04   3rd Qu.:13.00   3rd Qu.:135.64   3rd
Qu.:130.24
##  Max.   :2020-12-20   Max.   :28.00   Max.   :157.24   Max.
:152.76
##  walkingweekmean    stringency        retail          grocery
##  Min.   : 45.80   Min.   :11.11   Min.   :-65.14   Min.   :-30.0000
##  1st Qu.: 82.05   1st Qu.:55.09   1st Qu.:-41.86   1st Qu.: -6.0000
##  Median :111.76   Median :59.72   Median :-12.00   Median :  0.8571
##  Mean   :107.86   Mean   :57.82   Mean   :-22.82   Mean   : -2.4921
##  3rd Qu.:137.49   3rd Qu.:64.81   3rd Qu.: -5.00   3rd Qu.:  3.1429
##  Max.   :155.94   Max.   :82.41   Max.   :  5.50   Max.   : 27.0000
##      parks            transit         workplace       residential
##  Min.   :-16.71   Min.   :-58.00   Min.   :-58.57   Min.   : 0.000
##  1st Qu.:  8.00   1st Qu.:-39.71   1st Qu.:-25.57   1st Qu.: 3.143
##  Median : 43.57   Median :-21.43   Median :-17.43   Median : 5.143
##  Mean   : 51.32   Mean   :-26.49   Mean   :-21.64   Mean   : 6.902
##  3rd Qu.: 90.43   3rd Qu.:-16.29   3rd Qu.:-14.57   3rd Qu.:11.000
##  Max.   :149.43   Max.   :  2.00   Max.   :  0.00   Max.   :18.857

lapply(gon_and_activity[2:12], shapiro.test)

## $goncases
##
```

```
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.91635, p-value = 0.003187
##
##
## $drivingweekmean
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.94724, p-value = 0.03998
##
##
## $transitweekmean
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.93305, p-value = 0.01204
##
##
## $walkingweekmean
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.94737, p-value = 0.04044
##
##
## $stringency
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.86318, p-value = 8.142e-05
##
##
## $retail
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.8735, p-value = 0.0001564
##
##
## $grocery
##
##  Shapiro-Wilk normality test
##
```

```
## data:  X[[i]]
## W = 0.87986, p-value = 0.000237
##
##
## $parks
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.91429, p-value = 0.002722
##
##
## $transit
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.93845, p-value = 0.01887
##
##
## $workplace
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.91802, p-value = 0.003626
##
##
## $residential
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.90255, p-value = 0.001138
```

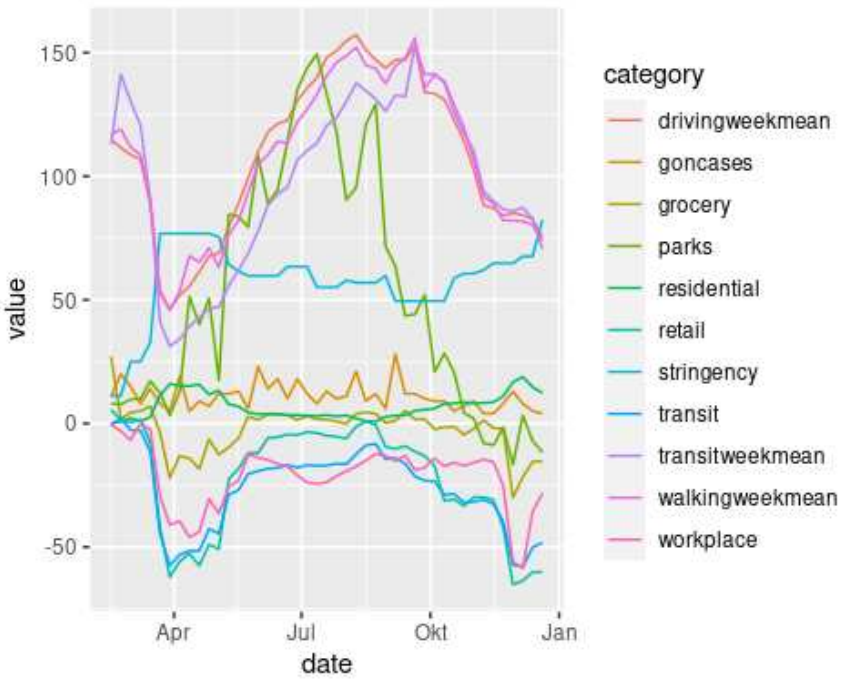
```
spearmanf<-function(x) {
  cor.test(gon_and_activity$goncases, x)$p.value
}
p<-as.data.frame(lapply(gon_and_activity[3:12], spearmanf))
p<-as.vector(p)
padj<-p.adjust(p,"holm",10)
padj
```

```
## drivingweekmean transitweekmean walkingweekmean      stringency
retail
##      0.22996525      0.34164990      0.26041078      0.22996525
0.05745214
##         grocery           parks         transit       workplace
residential
```

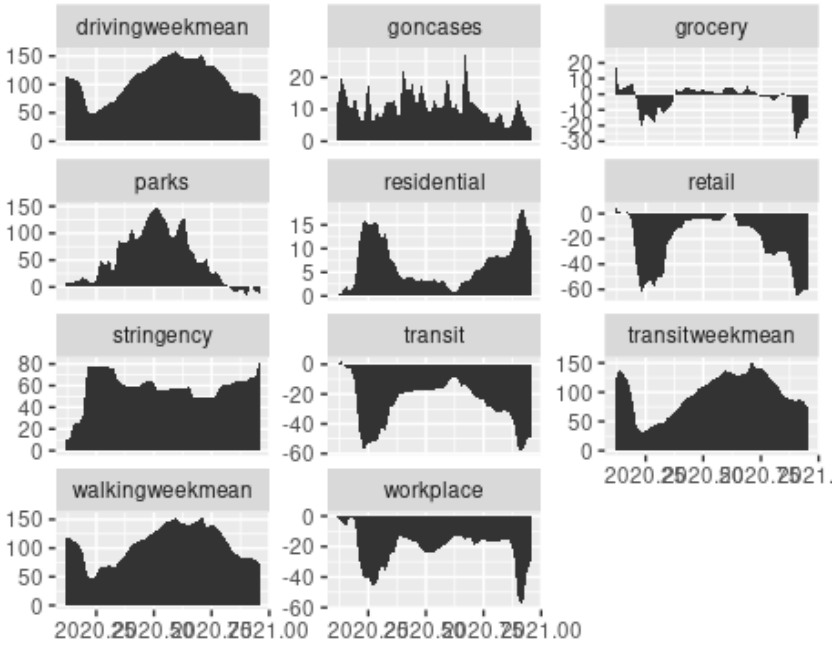
```
##      0.34164990      0.20367280      0.06505678      0.34164990
0.08303123

est<-as.data.frame(lapply(gon_and_activity[3:12], spearmanf))
est

##   drivingweekmean transitweekmean walkingweekmean stringency
retail
## 1        0.038622       0.1693287       0.0651027 0.03832754
0.005745214
##     grocery      parks     transit workplace residential
## 1 0.1138833 0.02909611 0.007228531 0.1732345   0.0103789

contemp<-rbind(est, padj)



library(Hmisc)

## Loading required package: lattice

## Loading required package: survival

## Loading required package: Formula

## Loading required package: ggplot2

##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:base':
##
##     format.pval, units

ccfspear<-function(y){ ccfspearmanx <- sapply( -3:3, function(l)
cor.test(y, Hmisc::Lag(gon_and_activity$goncases,l),method =
"spearman", use = "complete.obs", exact=FALSE)$estimate)
  }

est<-as.data.frame(lapply(gon_and_activity[3:12], ccfspear))

ccfspear<-function(y){ ccfspearmanx <- sapply( -3:3, function(l)
cor.test(y, Hmisc::Lag(gon_and_activity$goncases,l),method =
"spearman", use = "complete.obs", exact=FALSE)$p.value)
}

p<-as.data.frame(lapply(gon_and_activity[3:12], ccfspear))
p<-as.matrix(p)
padjust<-matrix(p.adjust(p, "holm", 70), nrow=7, ncol=10)
padjust
```

```
##          [,1] [,2] [,3]      [,4]      [,5]       [,6]       [,7]
[,8]
## [1,] 1.000000    1    1 1.0000000 1.0000000 1.00000000 0.07516131
1.00000000
## [2,] 1.000000    1    1 1.0000000 1.0000000 1.00000000 0.28852541
1.00000000
## [3,] 1.000000    1    1 1.0000000 0.2789515 1.00000000 1.00000000
0.27863624
## [4,] 0.929459    1    1 0.3776859 0.1061867 0.40303813 0.26705075
0.09989371
## [5,] 1.000000    1    1 1.0000000 0.0239737 0.02832612 0.18313202
0.18855380
## [6,] 1.000000    1    1 0.4588861 0.1746424 0.51717661 0.40303813
0.28739363
## [7,] 1.000000    1    1 1.0000000 0.3203195 0.22256502 0.47933840
0.26705075
##      [,9]     [,10]
## [1,]    1 1.0000000
## [2,]    1 1.0000000
## [3,]    1 0.2752604
## [4,]    1 0.1831320
## [5,]    1 0.1831320
## [6,]    1 0.2605047
## [7,]    1 0.1621512

est

##   drivingweekmean transitweekmean walkingweekmean  stringency
retail
## 1      0.09999678     -0.09641674      0.03457985 -0.04364172
0.2761435
## 2      0.16091375     -0.06874648      0.08057057 -0.04581780
0.3194382
## 3      0.24237221      0.08331324      0.19972545 -0.23447763
0.4135688
## 4      0.34166365      0.23987181      0.29962494 -0.39219850
0.4568015
## 5      0.32926902      0.24582025      0.29843541 -0.31520661
0.5154748
## 6      0.33218619      0.27035591      0.30566585 -0.38880019
0.4458953
## 7      0.24862941      0.20196058      0.21927845 -0.23784260
0.4140939
##     grocery     parks   transit workplace residential
## 1 0.1953446 0.4859078 0.2673252 0.1191726  -0.2686112
## 2 0.2986783 0.4149801 0.3079655 0.1519113  -0.3164139
## 3 0.3210346 0.3373550 0.4144321 0.2603637  -0.4157579
## 4 0.3882109 0.4141077 0.4596441 0.1960678  -0.4338489
## 5 0.5096071 0.4377524 0.4353386 0.2241331  -0.4380836
```

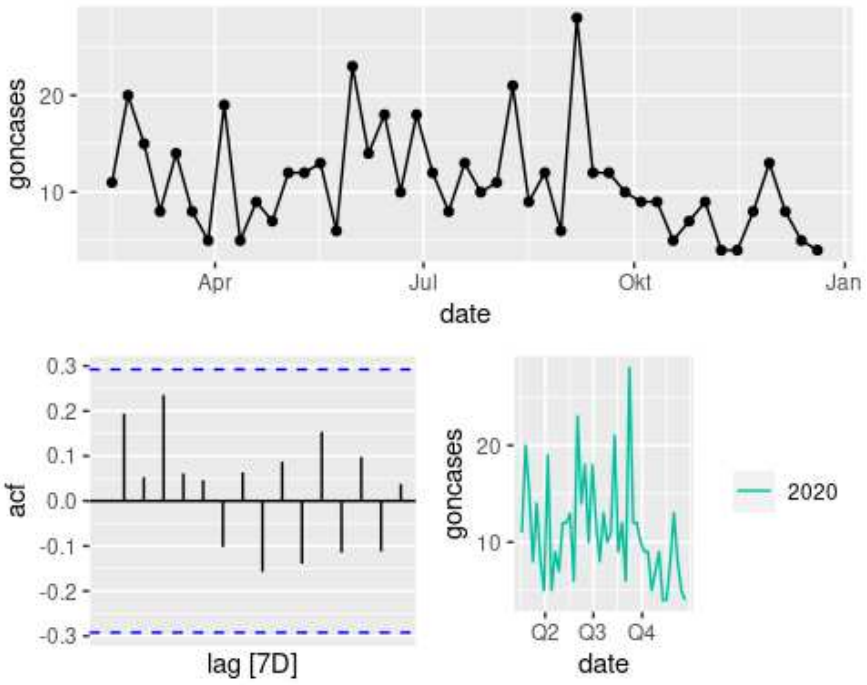
```
## 6 0.3808334 0.3969716 0.4160154 0.2638694  -0.4250761
## 7 0.4373978 0.3900994 0.4277721 0.2894795  -0.4545364

library(ggplot2)
library(tidyr)

gon_lon<-gather(gon_and_activity, category, value,
goncases:residential)
ggplot(gon_lon, aes(date, value, group=category, col=category)) +
geom_line()# + scale_y_log10()
```




```
####################modelling##############################################
library(ggfortify)
autoplot(ts(gon_and_activity[2:12], start=c(2020,7), frequency=52.18),
ncol=3, geom="ribbon")
```




```
library(tsibble)

##
## Attaching package: 'tsibble'

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

library(fable)

## Loading required package: fabletools

library(feasts)

gon<-as_tsibble(gon_and_activity)

## Using `date` as index variable.

gg_tsdisplay(gon) # looks stationary

## Plot variable not specified, automatically selected `y = goncases`
```




```
gon %>%
  features(goncases, unitroot_kpss) # and is statinary

## # A tibble: 1 x 2
##   kpss_stat kpss_pvalue
##       <dbl>       <dbl>
## 1     0.405      0.0751

fit <- gon %>%
  model(
    arima = ARIMA(goncases ~ drivingweekmean + transitweekmean +
walkingweekmean + stringency + retail + grocery + transit + parks +
workplace + residential)
  )
fit

## # A mable: 1 x 1
##                          arima
##                        <model>
## 1 <LM w/ ARIMA(1,0,0) errors>

glance(fit)

## # A tibble: 1 x 8
##   .model sigma2 log_lik   AIC  AICc   BIC ar_roots  ma_roots
##   <chr>   <dbl>   <dbl> <dbl> <dbl> <dbl> <list>    <list>
## 1 arima    20.7   -125.  276.  288.  300. <cpl [1]> <cpl [0]>
```

```
report(fit)
```

```
## Series: goncases
## Model: LM w/ ARIMA(1,0,0) errors
##
## Coefficients:
##           ar1  drivingweekmean  transitweekmean  walkingweekmean
stringency
##       -0.4765           0.2241          -0.2839           0.0712
-0.0401
## s.e.   0.1369           0.1623           0.0799           0.1506
0.1177
##       retail  grocery  transit    parks  workplace  residential
intercept
##       0.2237  -0.2240   0.4052  -0.0936    -0.2369       0.6133
20.0278
## s.e.  0.1709   0.1754   0.3429   0.0396     0.1959       1.3828
4.8608
##
## sigma^2 estimated as 20.71:  log likelihood=-125.19
## AIC=276.39   AICc=288.13   BIC=299.88
```

```
coefficients(fit)
```

```
## # A tibble: 12 x 6
##    .model term            estimate std.error statistic  p.value
##    <chr>  <chr>              <dbl>     <dbl>     <dbl>    <dbl>
##  1 arima  ar1               -0.476     0.137     -3.48  0.00113
##  2 arima  drivingweekmean    0.224     0.162      1.38  0.174
##  3 arima  transitweekmean   -0.284     0.0799    -3.55  0.000911
##  4 arima  walkingweekmean    0.0712    0.151      0.473 0.639
##  5 arima  stringency        -0.0401    0.118     -0.341 0.735
##  6 arima  retail             0.224     0.171      1.31  0.197
##  7 arima  grocery           -0.224     0.175     -1.28  0.208
##  8 arima  transit            0.405     0.343      1.18  0.244
##  9 arima  parks             -0.0936    0.0396    -2.36  0.0226
## 10 arima  workplace         -0.237     0.196     -1.21  0.233
## 11 arima  residential        0.613     1.38       0.443 0.660
## 12 arima  intercept         20.0       4.86       4.12  0.000160
```

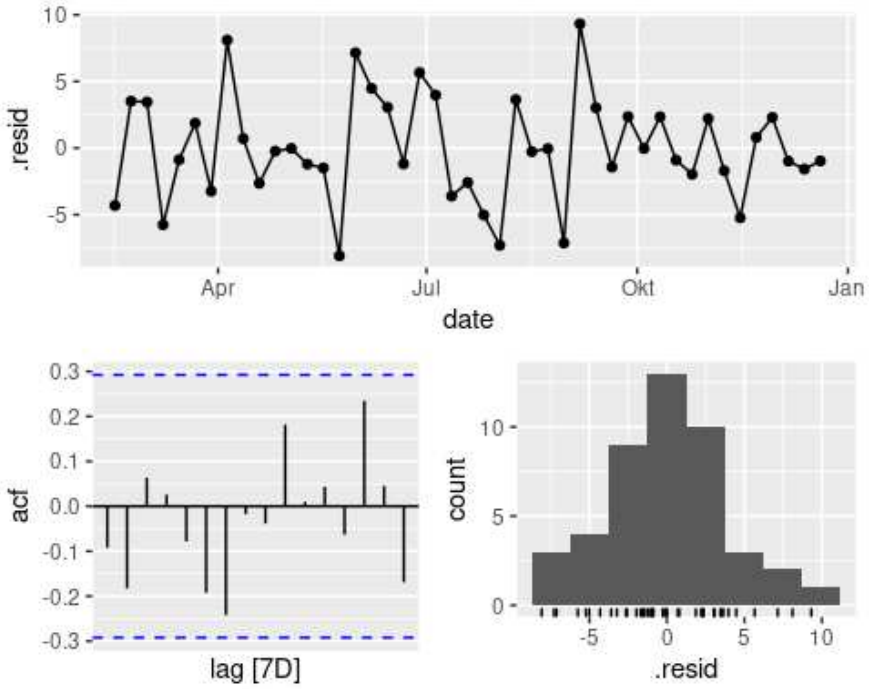
```
accuracy(fit)
```

```
## # A tibble: 1 x 10
##   .model .type         ME  RMSE   MAE   MPE  MAPE  MASE RMSSE
ACF1
##   <chr>  <chr>      <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
<dbl>
## 1 arima  Training -0.0393  3.90  3.06 -12.1  31.7 0.536 0.519 -
0.0922
```

```
report(fit)
```

```
## Series: goncases
## Model: LM w/ ARIMA(1,0,0) errors
##
## Coefficients:
##           ar1  drivingweekmean  transitweekmean  walkingweekmean
stringency
##       -0.4765           0.2241          -0.2839           0.0712
-0.0401
## s.e.   0.1369           0.1623           0.0799           0.1506
0.1177
##       retail  grocery  transit    parks  workplace  residential
intercept
##       0.2237  -0.2240   0.4052  -0.0936    -0.2369       0.6133
20.0278
## s.e.  0.1709   0.1754   0.3429   0.0396     0.1959       1.3828
4.8608
##
## sigma^2 estimated as 20.71:  log likelihood=-125.19
## AIC=276.39   AICc=288.13   BIC=299.88
```

```
gg_tsresiduals(fit) #looks fine
```




```
augment(fit) %>%
  features(.innov, ljung_box)

## # A tibble: 1 x 3
##   .model lb_stat lb_pvalue
```

```
##   <chr>    <dbl>     <dbl>
## 1 arima    0.409     0.523
```