주별 시계열을 통한 상관분석

```
load('../../refinedata/analysis/analysis_total_Fixed.rda')
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.6.3
```

```
##
## 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(FinCal)
```

```
## Warning: package 'FinCal' was built under R version 3.6.3
```

```
library(car)
```

```
## Warning: package 'car' was built under R version 3.6.3
```

```
## Loading required package: carData
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:dplyr':
##
##     recode
```

```
library(gvlma)
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.6.3
```

```
library(tidyr)
library(forecast)
```

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

```
n <- rep(1:157,each = 7)
analysis_total_Fixed$주 <- rep(n[1:1096], 17)
analysis_total_Fixed <- as.data.frame(analysis_total_Fixed)
analysis_total_week <- analysis_total_Fixed %>%
  group_by(주) %>%
  summarise(`평균기온(°C)` = mean(`평균기온(°C)`),
            `평균 풍속(m/s)` = mean(`평균 풍속(m/s)`),
            `평균 현지기압(hPa)` = mean(`평균 현지기압(hPa)`),
            `일강수량(mm)` = mean(`일강수량(mm)`),
            SO2 = geometric.mean(SO2),
            CO = geometric.mean(CO),
            O3 = geometric.mean(O3),
            NO2 = geometric.mean(NO2),
            PM10 = geometric.mean(PM10),
            PM25 = geometric.mean(PM25),
            발병률 = sum(발병률)
            )
```

```
#다중 공산성 확인
fit <- lm(data = analysis_total_week[c(-1,-2,-11)], formula = 발병률 ~ .)
summary(fit)
```

```
##
## Call:
## lm(formula = 발병률 ~ ., data = analysis_total_week[c(-1, -2,
##     -11)])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.07218 -0.19347 -0.02517  0.23217  1.12918
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -2.103e+01  8.559e+00  -2.457   0.0152 *
## `평균 풍속(m/s)`      4.070e-01  9.392e-02   4.333 2.70e-05 ***
## `평균 현지기압(hPa)`  2.118e-02  8.623e-03   2.456   0.0152 *
## `일강수량(mm)`       -4.545e-03  7.408e-03  -0.614   0.5404
## SO2                  -3.586e+02  8.945e+01  -4.009 9.64e-05 ***
## CO                    1.113e+00  1.048e+00   1.062   0.2900
## O3                    7.153e+00  6.250e+00   1.145   0.2542
## NO2                   9.501e+01  2.260e+01   4.204 4.51e-05 ***
## PM10                  3.276e-03  5.114e-03   0.641   0.5228
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3604 on 148 degrees of freedom
## Multiple R-squared:  0.6799, Adjusted R-squared:  0.6626
## F-statistic: 39.29 on 8 and 148 DF,  p-value: < 2.2e-16
```

```
vif(fit)
```

```
##     `평균 풍속(m/s)` `평균 현지기압(hPa)`       `일강수량(mm)`
##             1.501922             4.141342             1.537544
##                  SO2                   CO                   O3
##             2.812274             9.152586             3.473381
##                  NO2                 PM10
##            11.355846             4.041523
```

```
fit <- lm(data = analysis_total_week[-1], formula = 발병률 ~ SO2 + CO + O3 + NO2 + PM10)
summary(fit)
```

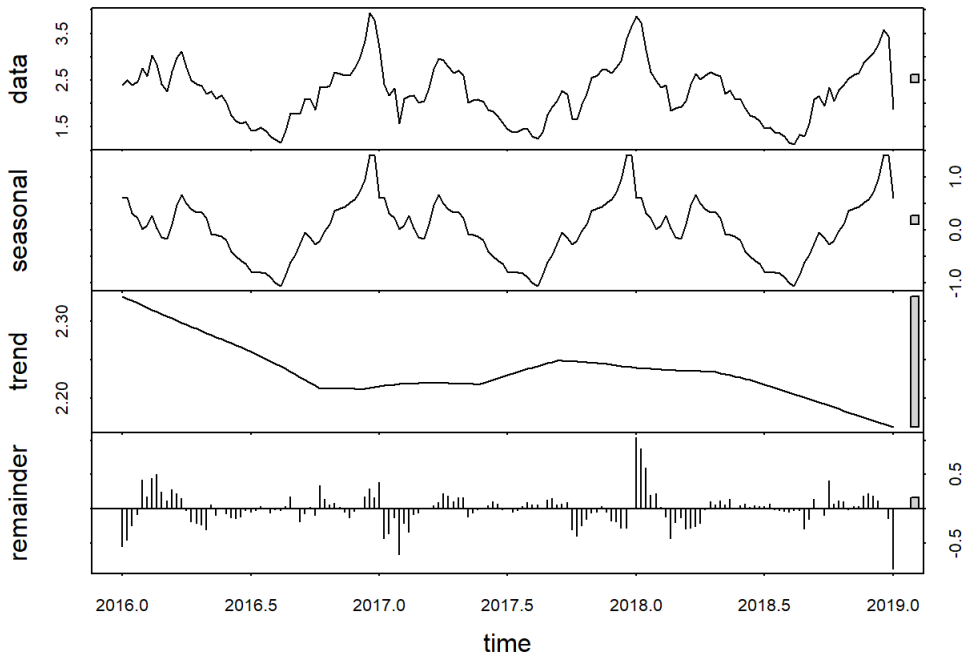
```
##
## Call:
## lm(formula = 발병률 ~ SO2 + CO + O3 + NO2 + PM10, data = analysis_total_week[-1])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.03477 -0.26660 -0.06029  0.28903  1.28927
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.358e-01  3.138e-01   1.389 0.166903
## SO2         -2.852e+02  9.866e+01  -2.891 0.004409 **
## CO           2.590e+00  1.162e+00   2.229 0.027309 *
## O3           6.736e+00  6.373e+00   1.057 0.292188
## NO2          8.965e+01  2.357e+01   3.803 0.000207 ***
## PM10         6.046e-04  5.729e-03   0.106 0.916093
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4085 on 151 degrees of freedom
## Multiple R-squared:  0.5803, Adjusted R-squared:  0.5664
## F-statistic: 41.76 on 5 and 151 DF,  p-value: < 2.2e-16
```

```
vif(fit)
```

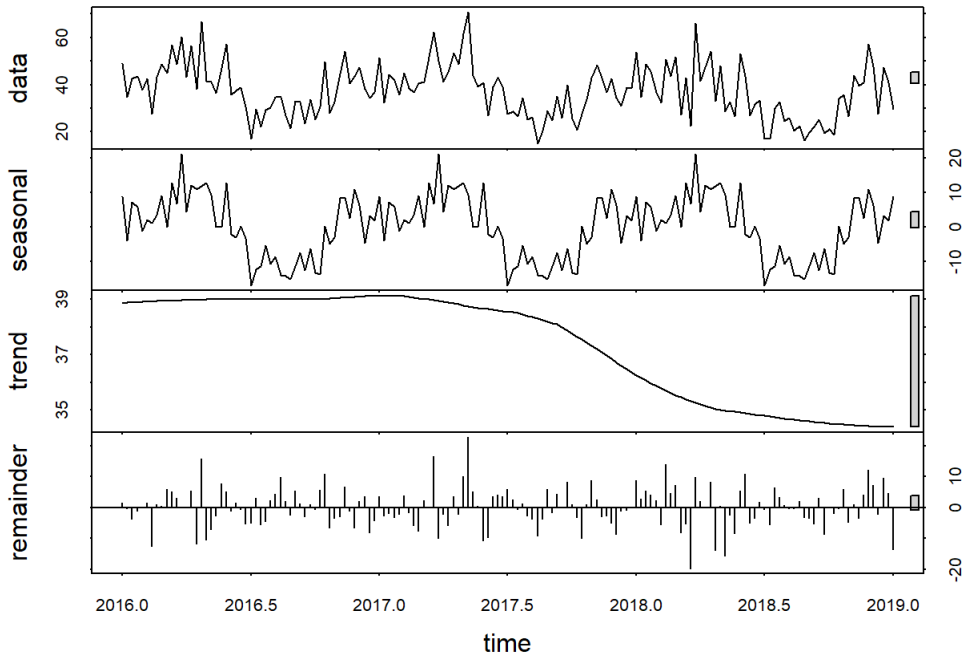
```
##      SO2       CO       O3      NO2     PM10
## 2.662578 8.751625 2.810290 9.614616 3.946582
```

```
# 주별 시계열 그래프 및 시계열 자료형을 통한 회귀분석
analysis_total_week <- as.data.frame(analysis_total_week)
ts_week <- ts(analysis_total_week[-1])
```

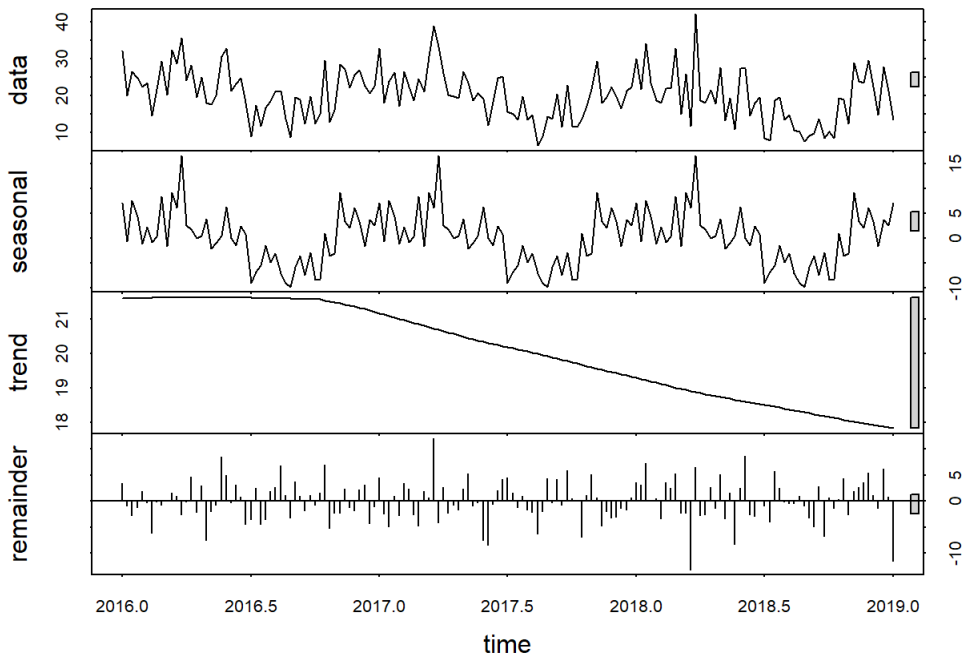
```
# 발병률
ts <- ts(analysis_total_week$발병률, start = c(2016, 1), freq = 52)

fit <- stl(ts, s.window = 'periodic')
plot(fit)
```



```
# 미세먼지

ts <- ts(analysis_total_week$PM10, start = c(2016, 1), freq = 52)

fit <- stl(ts, s.window = 'periodic')
plot(fit)
```

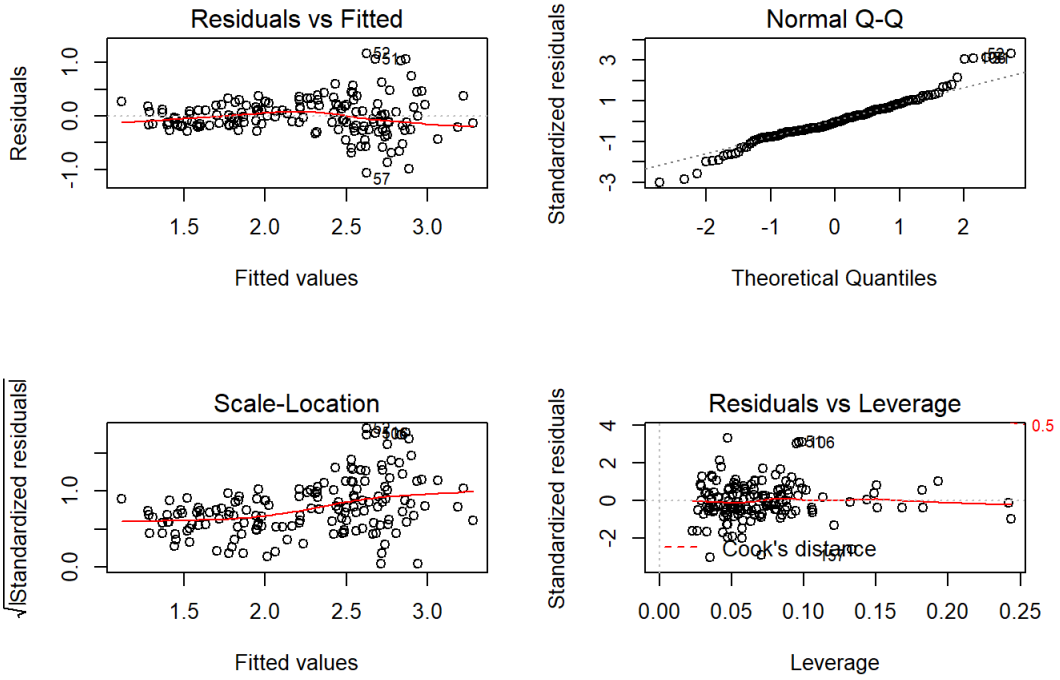



```
# PM25

ts <- ts(analysis_total_week$PM25, start = c(2016, 1), freq = 52)

fit <- stl(ts, s.window = 'periodic')
plot(fit)
```



```
ts<-ts(analysis_total_week[-1], start = c(2016, 1), freq = 52)
ts_temperature <- ts(analysis_total_week$`평균기온(°C)`, start = c(2016, 1), freq = 52)
ts_air <- ts(analysis_total_week$`평균 현지기압(hPa)`, start = c(2016, 1), freq = 52)
ts_wind <- ts(analysis_total_week$`평균 풍속(m/s)`, start = c(2016, 1), freq = 52)
ts_rain <- ts(analysis_total_week$`일강수량(mm)`, start = c(2016, 1), freq = 52)
ts_CO <- ts(analysis_total_week$CO, start = c(2016, 1), freq = 52)
ts_발병률 <- ts(analysis_total_week$발병률, start = c(2016, 1), freq = 52)
ts_NO2 <- ts(analysis_total_week$NO2, start = c(2016, 1), freq = 52)
ts_SO2 <- ts(analysis_total_week$SO2, start = c(2016, 1), freq = 52)
ts_O3 <- ts(analysis_total_week$O3, start = c(2016, 1), freq = 52)
ts_PM10 <- ts(analysis_total_week$PM10, start = c(2016, 1), freq = 52)
ts_PM25 <- ts(analysis_total_week$PM25, start = c(2016, 1), freq = 52)
```

```
fit <- lm(ts_발병률 ~ ts_temperature + ts_air + ts_rain + ts_wind + ts_CO + ts_SO2 + ts_NO2 + ts_O3 + ts_PM10
+ ts_PM25)
par(mfrow = c(2,2))
plot(fit)
```




```
summary(fit)
```

```
##
## Call:
## lm(formula = ts_발병률 ~ ts_temperature + ts_air + ts_rain +
##     ts_wind + ts_CO + ts_SO2 + ts_NO2 + ts_O3 + ts_PM10 + ts_PM25)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.06260 -0.18129 -0.02386  0.19600  1.16481
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -1.477e+01  1.277e+01  -1.157 0.249354
## ts_temperature -7.472e-03  1.255e-02  -0.595 0.552622
## ts_air          1.520e-02  1.229e-02   1.237 0.218046
## ts_rain        -3.316e-03  7.508e-03  -0.442 0.659430
## ts_wind         2.736e-01  1.491e-01   1.834 0.068619 .
## ts_CO           1.479e+00  1.214e+00   1.218 0.225208
## ts_SO2         -3.387e+02  9.300e+01  -3.642 0.000376 ***
## ts_NO2          9.310e+01  2.351e+01   3.959 0.000117 ***
## ts_O3           9.790e+00  6.515e+00   1.503 0.135066
## ts_PM10         9.134e-03  6.418e-03   1.423 0.156791
## ts_PM25        -1.720e-02  1.332e-02  -1.291 0.198845
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3598 on 146 degrees of freedom
## Multiple R-squared:  0.6853, Adjusted R-squared:  0.6637
## F-statistic: 31.79 on 10 and 146 DF,  p-value: < 2.2e-16
```

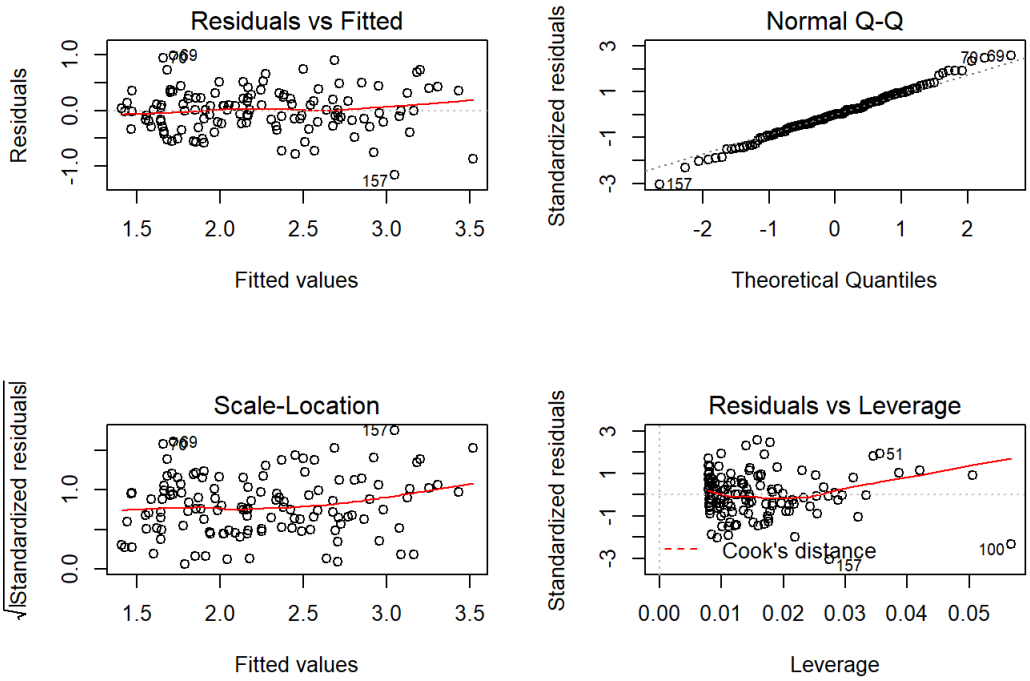
```
gvlma(fit)
```

```
##
## Call:
## lm(formula = ts_발병률 ~ ts_temperature + ts_air + ts_rain +
##     ts_wind + ts_CO + ts_SO2 + ts_NO2 + ts_O3 + ts_PM10 + ts_PM25)
##
## Coefficients:
##    (Intercept)  ts_temperature          ts_air         ts_rain         ts_wind
##     -1.477e+01      -7.472e-03       1.520e-02      -3.316e-03       2.736e-01
##          ts_CO          ts_SO2          ts_NO2           ts_O3         ts_PM10
##      1.479e+00      -3.387e+02       9.310e+01       9.790e+00       9.134e-03
##        ts_PM25
##     -1.720e-02
##
##
## ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS
## USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:
## Level of Significance =  0.05
##
## Call:
##  gvlma(x = fit)
##
##                      Value   p-value                   Decision
## Global Stat        25.8664 3.367e-05 Assumptions NOT satisfied!
## Skewness            2.2255 1.358e-01    Assumptions acceptable.
## Kurtosis           21.3912 3.745e-06 Assumptions NOT satisfied!
## Link Function       1.9750 1.599e-01    Assumptions acceptable.
## Heteroscedasticity  0.2747 6.002e-01    Assumptions acceptable.
```

```
vif(fit)
```

```
## ts_temperature         ts_air        ts_rain        ts_wind          ts_CO
##      17.813074       8.438324       1.584735       3.799551      12.322635
##         ts_SO2         ts_NO2          ts_O3        ts_PM10        ts_PM25
##       3.049735      12.335508       3.786624       6.386078      10.106251
```

```
ts_O3 <- ts(lag(analysis_total_week$O3,30), start = c(2016, 1), freq = 52)
fit <- lm( ts_발병률~ts_O3)
par(mfrow = c(2,2))
plot(fit)
```



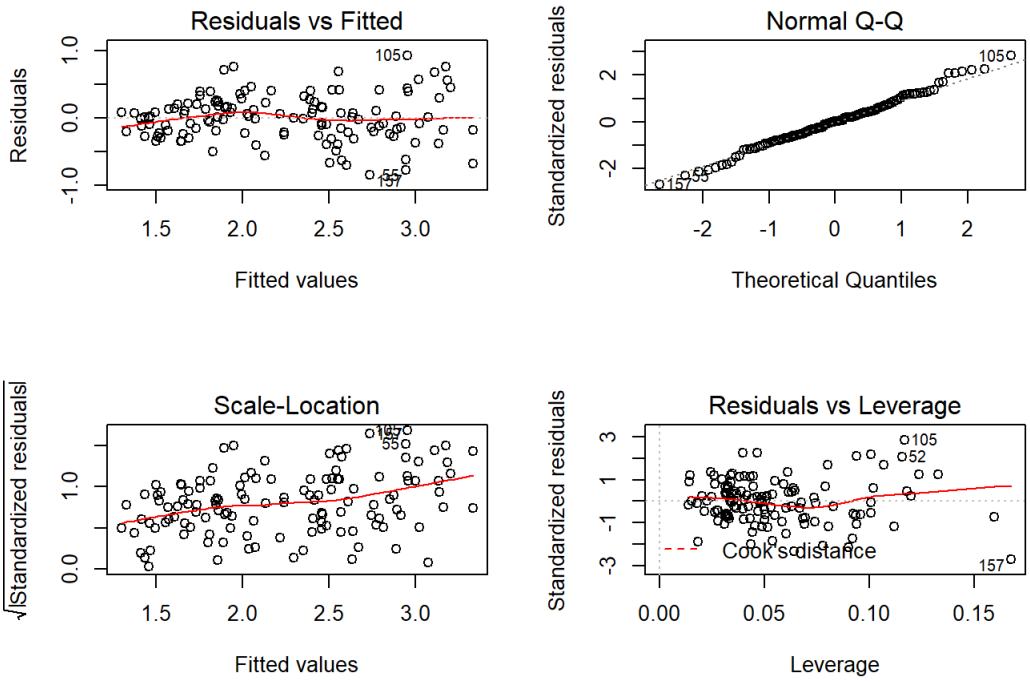

```
summary(fit)
```

```
##
## Call:
## lm(formula = ts_발병률 ~ ts_O3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.16521 -0.21479  0.00349  0.22954  0.98999
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.7437     0.1053   7.064 9.96e-11 ***
## ts_O3        59.3144     3.9561  14.993  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3871 on 125 degrees of freedom
##   (30 observations deleted due to missingness)
## Multiple R-squared:  0.6426, Adjusted R-squared:  0.6398
## F-statistic: 224.8 on 1 and 125 DF,  p-value: < 2.2e-16
```

```
gvlma(fit)
```

```
##
## Call:
## lm(formula = ts_발병률 ~ ts_O3)
##
## Coefficients:
## (Intercept)        ts_O3
##      0.7437      59.3144
##
##
## ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS
## USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:
## Level of Significance =  0.05
##
## Call:
##  gvlma(x = fit)
##
##                       Value p-value                Decision
## Global Stat        1.332705  0.8558 Assumptions acceptable.
## Skewness           0.004584  0.9460 Assumptions acceptable.
## Kurtosis           0.459367  0.4979 Assumptions acceptable.
## Link Function      0.019039  0.8903 Assumptions acceptable.
## Heteroscedasticity 0.849715  0.3566 Assumptions acceptable.
```

```
fit <- lm(formula = ts_발병률 ~ ts_air + ts_wind + ts_CO + ts_SO2 +
            ts_NO2 + ts_O3)
par(mfrow = c(2,2))
plot(fit)
```




```
summary(fit)
```

```
##
## Call:
## lm(formula = ts_발병률 ~ ts_air + ts_wind + ts_CO + ts_SO2 +
##     ts_NO2 + ts_O3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.85156 -0.21937  0.00197  0.20321  0.92313
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   5.852142   8.622036   0.679 0.498608
## ts_air       -0.006031   0.008752  -0.689 0.492038
## ts_wind       0.375702   0.095511   3.934 0.000141 ***
## ts_CO        -0.617374   1.173506  -0.526 0.599794
## ts_SO2      -99.672598 106.062398  -0.940 0.349231
## ts_NO2       82.446250  23.805290   3.463 0.000740 ***
## ts_O3        37.394183   6.973949   5.362 4.05e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3465 on 120 degrees of freedom
##   (30 observations deleted due to missingness)
## Multiple R-squared:  0.7252, Adjusted R-squared:  0.7115
## F-statistic: 52.79 on 6 and 120 DF,  p-value: < 2.2e-16
```

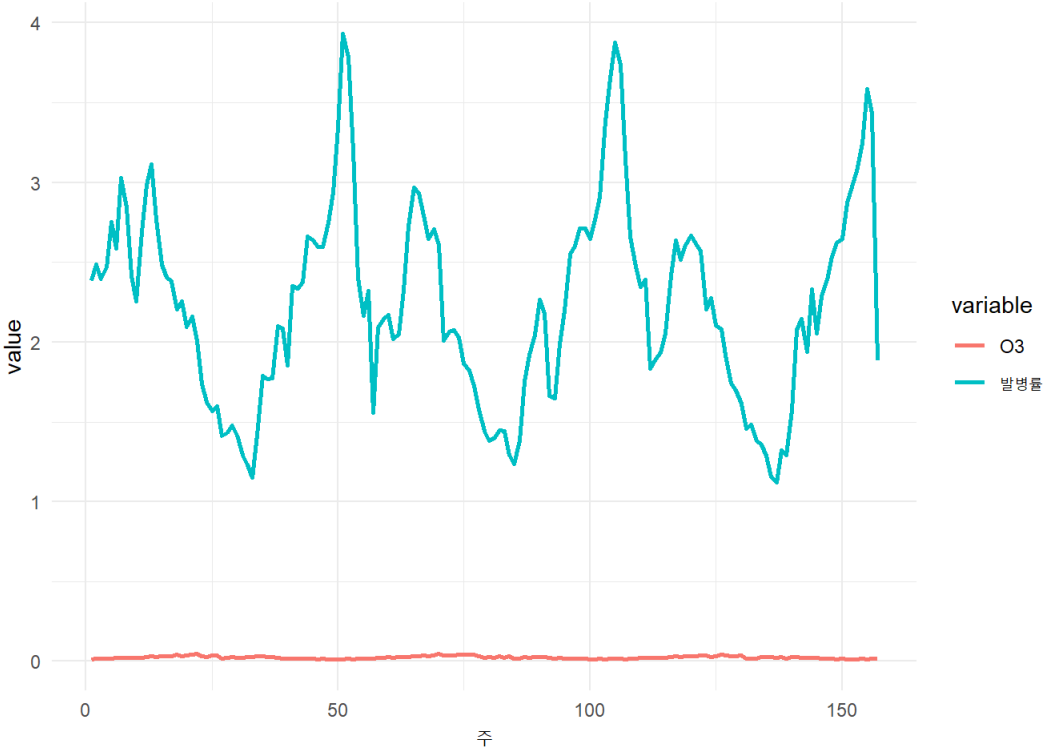
```
gvlma(fit)
```

```
##
## Call:
## lm(formula = ts_발병률 ~ ts_air + ts_wind + ts_CO + ts_SO2 +
##     ts_NO2 + ts_O3)
##
## Coefficients:
## (Intercept)       ts_air      ts_wind        ts_CO       ts_SO2       ts_NO2
##    5.852142    -0.006031     0.375702    -0.617374   -99.672598    82.446250
##       ts_O3
##   37.394183
##
##
## ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS
## USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:
## Level of Significance =  0.05
##
## Call:
##  gvlma(x = fit)
##
##                      Value p-value                Decision
## Global Stat        0.41595  0.9812 Assumptions acceptable.
## Skewness           0.11247  0.7373 Assumptions acceptable.
## Kurtosis           0.05877  0.8084 Assumptions acceptable.
## Link Function      0.22180  0.6377 Assumptions acceptable.
## Heteroscedasticity 0.02291  0.8797 Assumptions acceptable.
```

```
vif(fit)
```

```
##    ts_air   ts_wind     ts_CO    ts_SO2    ts_NO2     ts_O3
##  3.665507  1.357798 10.456317  2.960768 12.087291  3.879894
```

```
df <- analysis_total_week %>%
  dplyr::select(주, 발병률 ,O3) %>%
  gather(key = "variable", value = "value", -주)

ggplot(df, aes(x = 주, y = value)) +
  geom_line(aes(color = variable), size = 1) +
  # scale_color_manual(values = c("#00AFBB", "#E7B800", )) +
  theme_minimal()
```



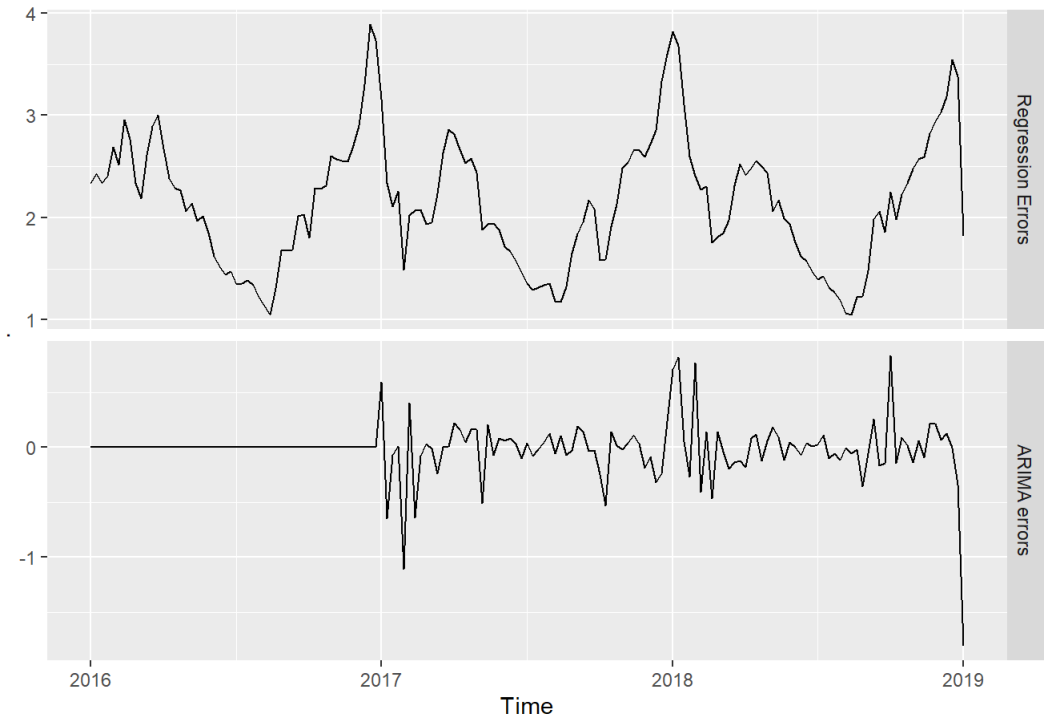

```
fit <- lm(발병률 ~ ., data = analysis_total_week[c(-1,-6,-8)])
summary(fit)
```

```
##
## Call:
## lm(formula = 발병률 ~ ., data = analysis_total_week[c(-1, -6,
##     -8)])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.11712 -0.19813 -0.03341  0.20019  1.18106
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)           -1.993e+01  1.237e+01  -1.612 0.109176
## `평균기온(°C)`        -9.028e-05  1.287e-02  -0.007 0.994411
## `평균 풍속(m/s)`       2.822e-01  1.488e-01   1.896 0.059908 .
## `평균 현지기압(hPa)`   1.990e-02  1.194e-02   1.667 0.097703 .
## `일강수량(mm)`        -1.838e-03  7.605e-03  -0.242 0.809324
## CO                     5.965e-01  1.191e+00   0.501 0.617152
## NO2                    8.385e+01  2.420e+01   3.464 0.000696 ***
## PM10                   1.164e-02  6.478e-03   1.796 0.074507 .
## PM25                  -2.460e-02  1.273e-02  -1.933 0.055129 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3748 on 148 degrees of freedom
## Multiple R-squared:  0.6538, Adjusted R-squared:  0.6351
## F-statistic: 34.93 on 8 and 148 DF,  p-value: < 2.2e-16
```

```
car::vif(fit)
```

```
##       `평균기온(°C)`      `평균 풍속(m/s)` `평균 현지기압(hPa)`
##            17.241488             3.487039             7.338034
##       `일강수량(mm)`                   CO                  NO2
##             1.498210            10.917786            12.044650
##                 PM10                 PM25
##             5.994910             8.495749
```

```
fit <- auto.arima(ts[,'발병률'],
                  xreg=ts[,'O3'])

cbind("Regression Errors" = residuals(fit, type="regression"),
      "ARIMA errors" = residuals(fit, type="innovation")) %>%
  autoplot(facets=TRUE)
```




```
checkresiduals(fit)
```

```
##
##  Ljung-Box test
##
## data:  Residuals from Regression with ARIMA(2,0,0)(0,1,0)[52] errors
## Q* = 22.977, df = 28, p-value = 0.7342
##
## Model df: 3.   Total lags used: 31
```

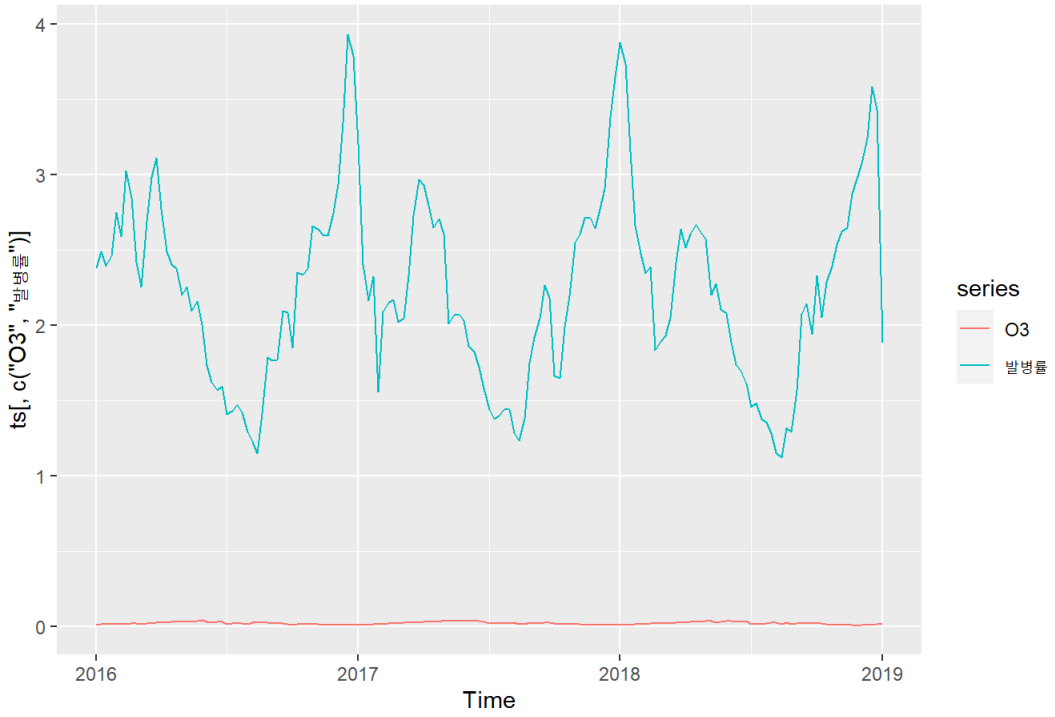
```
summary(fit)
```

```
## Series: ts[, "발병률"]
## Regression with ARIMA(2,0,0)(0,1,0)[52] errors
##
## Coefficients:
##          ar1      ar2    xreg
##       0.7575  -0.1091  3.5319
## s.e.  0.1227   0.1222  6.0773
##
## sigma^2 estimated as 0.1067:  log likelihood=-30.31
## AIC=68.63   AICc=69.03   BIC=79.25
##
## Training set error measures:
##                       ME      RMSE       MAE       MPE     MAPE      MASE
## Training set -0.01412122 0.2632992 0.1277896 -1.345115 5.901806 0.4905129
##                     ACF1
## Training set -0.04434632
```

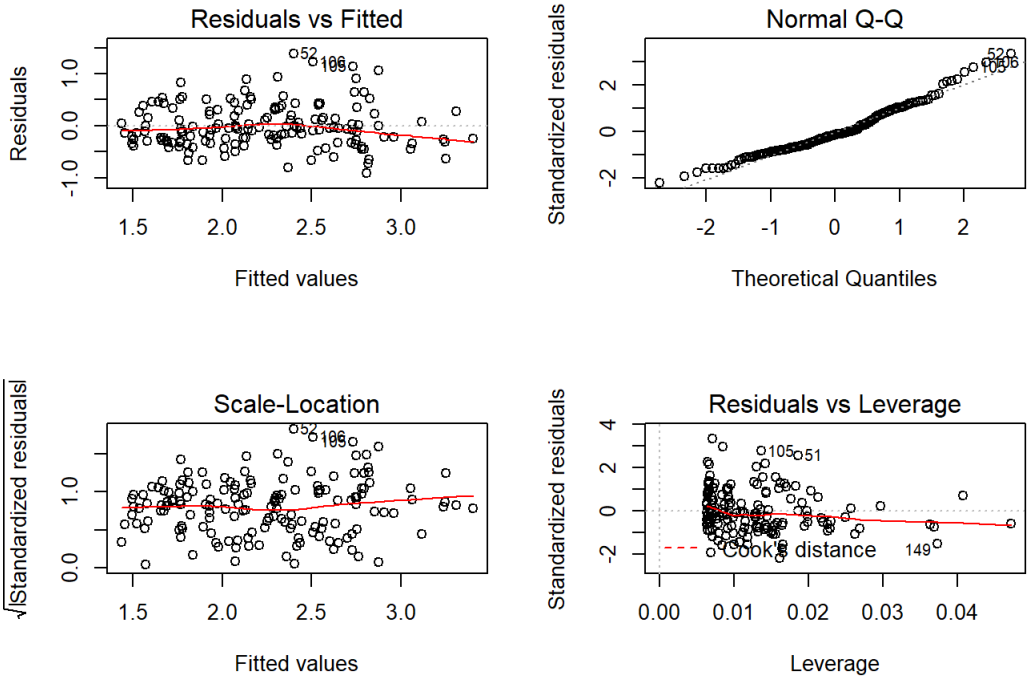
```
cor(analysis_total_week[c(-1,-2)])
```

```
##                    평균 풍속(m/s) 평균 현지기압(hPa) 일강수량(mm)          SO2
## 평균 풍속(m/s)          1.00000000          0.1104871  0.064791052  0.114317221
## 평균 현지기압(hPa)      0.11048706          1.0000000 -0.390418964  0.426401558
## 일강수량(mm)            0.06479105         -0.3904190  1.000000000 -0.449022274
## SO2                     0.11431722          0.4264016 -0.449022274  1.000000000
## CO                     -0.03742704          0.7473116 -0.428422544  0.697771856
## O3                      0.21039002         -0.6044263 -0.002005014 -0.007983935
## NO2                    -0.12355493          0.7501345 -0.441027555  0.680358259
## PM10                   -0.01192844          0.3506930 -0.479039517  0.691190509
## PM25                   -0.18414374          0.3946605 -0.471269757  0.736945407
## 발병률                  0.16778191          0.7072087 -0.353241861  0.423731392
##                              CO          O3         NO2         PM10         PM25
## 평균 풍속(m/s)      -0.03742704  0.210390017 -0.1235549 -0.01192844 -0.18414374
## 평균 현지기압(hPa)   0.74731159 -0.604426254  0.7501345  0.35069303  0.39466049
## 일강수량(mm)        -0.42842254 -0.002005014 -0.4410276 -0.47903952 -0.47126976
## SO2                   0.69777186 -0.007983935  0.6803583  0.69119051  0.73694541
## CO                    1.00000000 -0.456480844  0.9300881  0.63950077  0.74939661
## O3                   -0.45648084  1.000000000 -0.4335284  0.16043043  0.02834337
## NO2                   0.93008807 -0.433528432  1.0000000  0.69000099  0.76993968
## PM10                  0.63950077  0.160430431  0.6900010  1.00000000  0.88692765
## PM25                  0.74939661  0.028343374  0.7699397  0.88692765  1.00000000
## 발병률                0.71673585 -0.331021424  0.7429323  0.50541587  0.48583468
##                        발병률
## 평균 풍속(m/s)      0.1677819
## 평균 현지기압(hPa)  0.7072087
## 일강수량(mm)       -0.3532419
## SO2                 0.4237314
## CO                  0.7167359
## O3                 -0.3310214
## NO2                 0.7429323
## PM10                0.5054159
## PM25                0.4858347
## 발병률              1.0000000
```

```
autoplot(ts[,c('O3','발병률')])
```




```
fit <- lm(발병률 ~ NO2, data = analysis_total_week[c(-1,-2)])
par(mfrow = c(2,2))
plot(fit)
```




```
summary(fit)
```

```
##
## Call:
## lm(formula = 발병률 ~ NO2, data = analysis_total_week[c(-1, -2)])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.91162 -0.29070 -0.06279  0.27780  1.38524
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.4903     0.1308   3.749  0.00025 ***
## NO2         107.1248     7.7524  13.818  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4166 on 155 degrees of freedom
## Multiple R-squared:  0.5519, Adjusted R-squared:  0.5491
## F-statistic: 190.9 on 1 and 155 DF,  p-value: < 2.2e-16
```