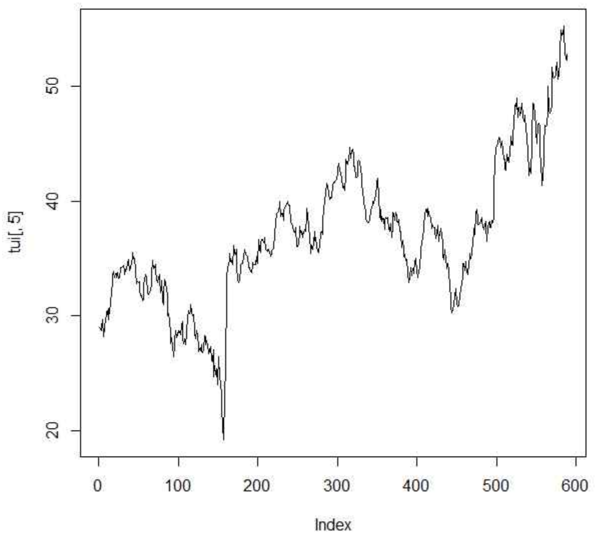
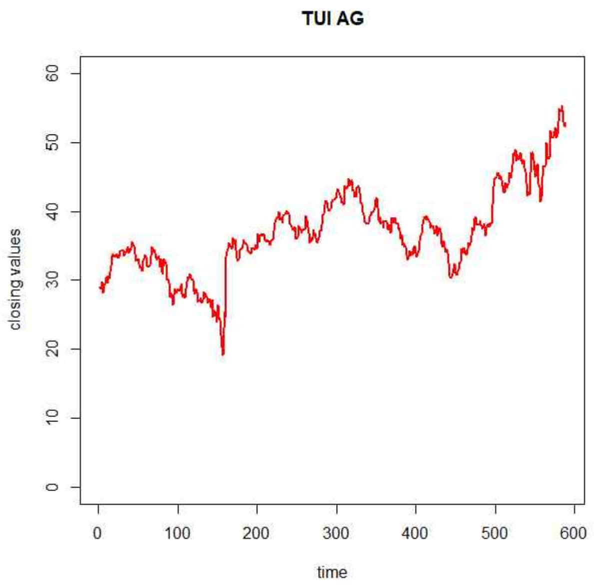
데이터 사이언스 과제8

< 1. Getting started >

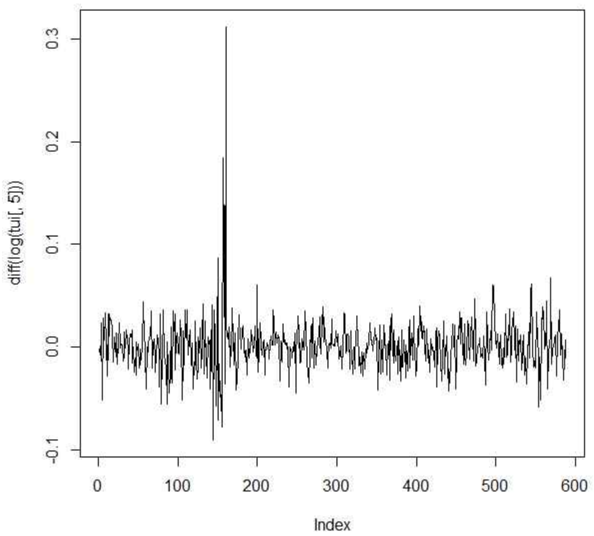
```
> tui <- read.csv("C:/Users/김황규/Documents/tui.csv", header=T, dec=",",
sep=";")
>
> plot(tui[,5],type="l")
```




```
> plot(tui[,5], type="l",
+  lwd=2, col="red", xlab="time", ylab="closing values",
+  main="TUI AG", ylim=c(0,60) )
```

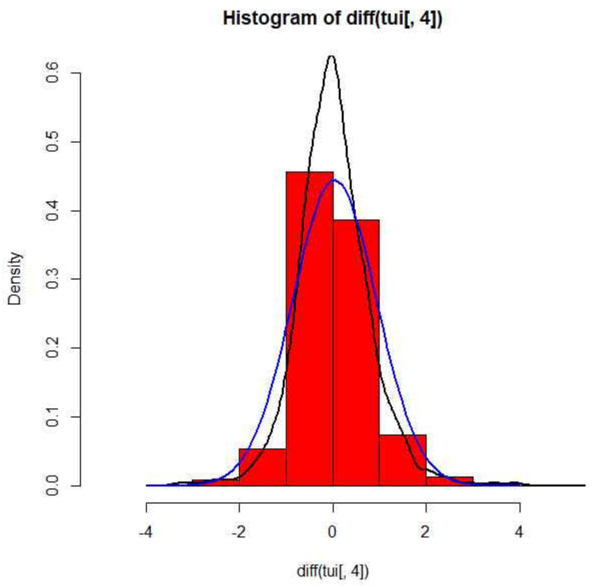



```
> plot(diff(log(tui[,5])),type="l")
```

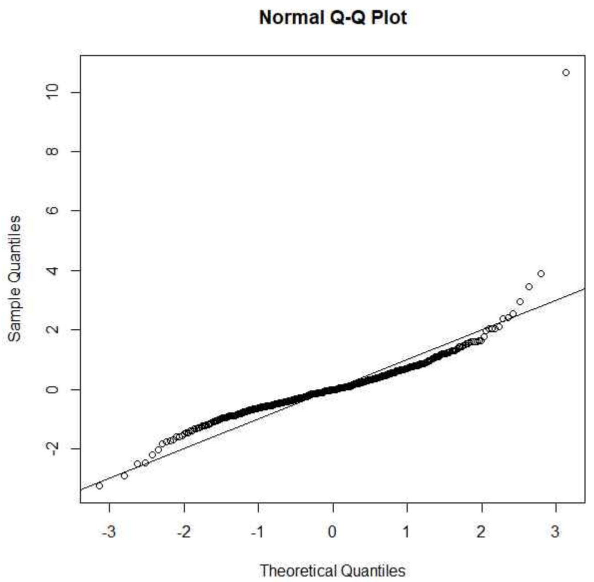



```
> hist(diff(tui[,4]),prob=T,ylim=c(0,0.6),xlim=c(-5,5),col="red")
> lines(density(diff(tui[,4])),lwd=2)
>
> mu<-mean(diff(tui[,4]))
> sigma<-sd(diff(tui[,4]))
>
```

```
> x<-seq(-4,4,length=100)
> y<-dnorm(x,mu,sigma)
> lines(x,y,lwd=2,col="blue")
```




```
> qqnorm(diff(tui[,4]))
> abline(0,1)
```

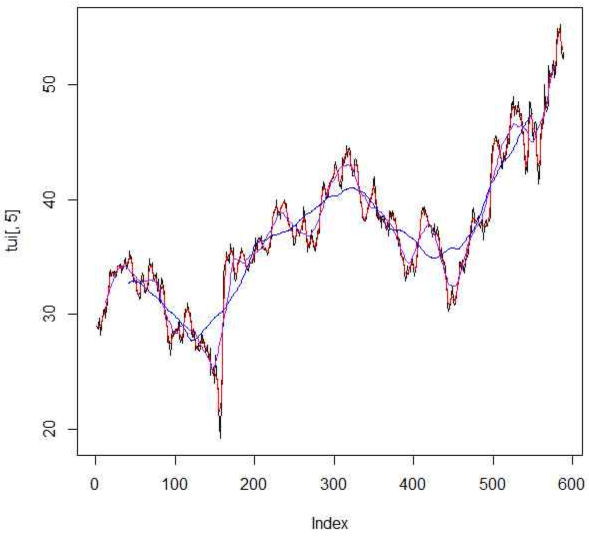



```
> x<-diff(log(tui[,5]))
```

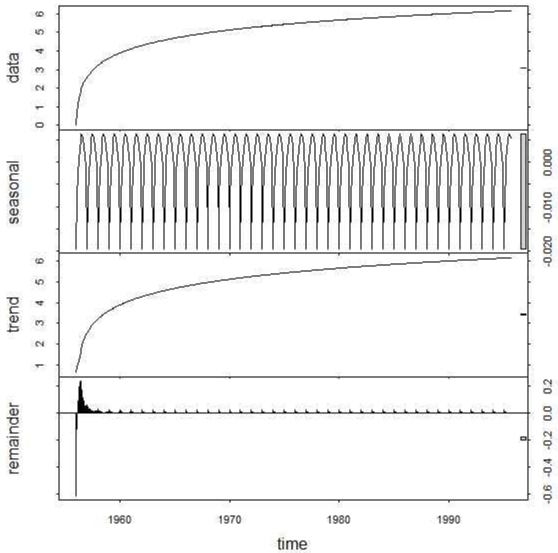
```
> ks.test(x,"pnorm",mean(x),sd(x))

        One-sample Kolmogorov-Smirnov test

data:  x
D = 0.098003, p-value = 2.487e-05
alternative hypothesis: two-sided

경고메시지(들):
In ks.test(x, "pnorm", mean(x), sd(x)) :
  Kolmogorov-Smirnov 테스트를 이용할 때는 ties가 있으면 안됩니다
>
> shapiro.test(x)

        Shapiro-Wilk normality test

data:  x
W = 0.80962, p-value < 2.2e-16
```

< 2. Simple Component Analysis >

```
> #library(ts)
> #install.packages("timeSeries") 로 바뀜
> library(timeSeries)
필요한 패키지를 로딩중입니다: timeDate
>
> plot(tui[,5],type="l")
>  tui.1 <- filter(tui[,5],filter=rep(1/5,5))
>  tui.2 <- filter(tui[,5],filter=rep(1/25,25))
>  tui.3 <- filter(tui[,5],filter=rep(1/81,81))
> lines(tui.1,col="red")
> lines(tui.2,col="purple")
> lines(tui.3,col="blue")
```

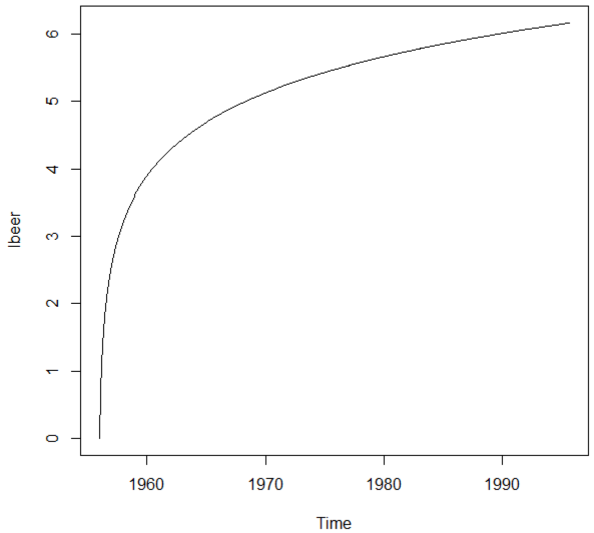



```
> beer<-read.csv("beer.csv",header=T,dec=",",sep=";")
> beer<-ts(beer[,1],start=1956,freq=12)
> plot(stl(log(beer),s.window="periodic"))
```

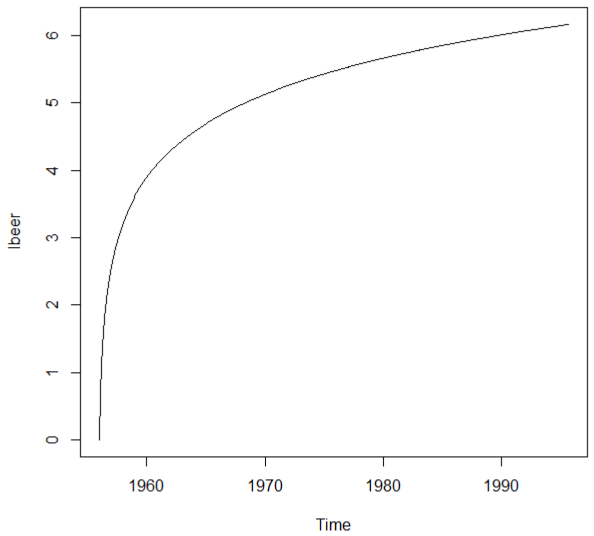



```
> lbeer<-log(beer)
> t<-seq(1956,1995.2,length=length(beer))
> t2<-t^2
> plot(lbeer)
```

```
> lm(lbeer~t+t2)

Call:
lm(formula = lbeer ~ t + t2)

Coefficients:
(Intercept)            t           t2
 -1.239e+04    1.247e+01   -3.138e-03

> lines(lm(lbeer~t+t2)$fit,col=2,lwd=2)
```



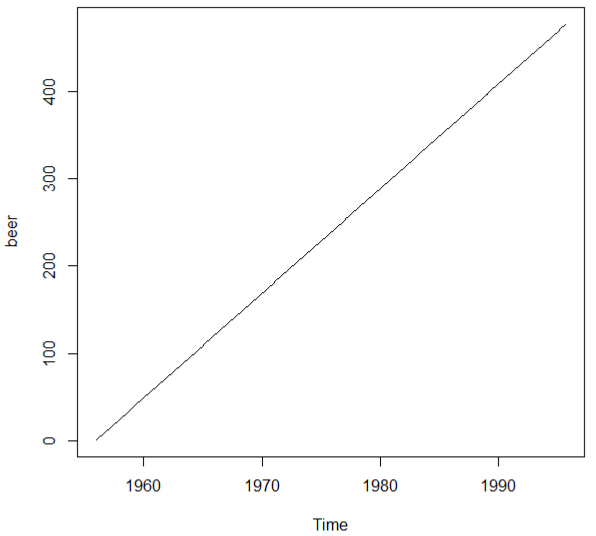

```
> lbeer<-log(beer)
> t<-seq(1956,1995.2,length=length(beer))
> t2<-t^2
> sin.t<-sin(2*pi*t)
> cos.t<-cos(2*pi*t)
> plot(lbeer)
> lines(lm(lbeer~t+t2+sin.t+cos.t)$fit,col=4)
```



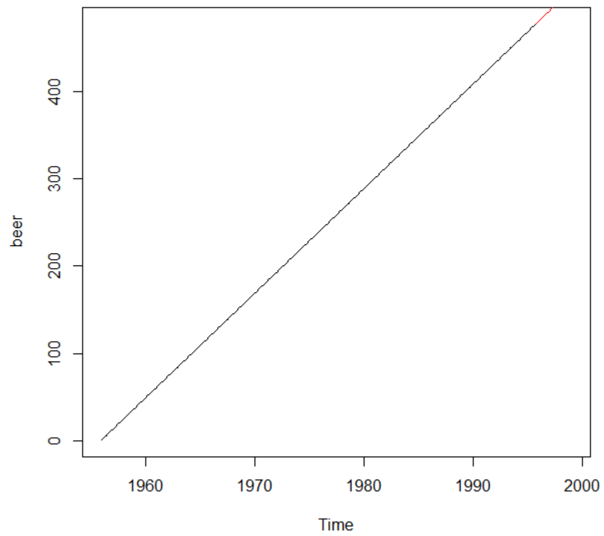

```
> summary(lm(lbeer~t+t2+sin.t+cos.t))

Call:
lm(formula = lbeer ~ t + t2 + sin.t + cos.t)

Residuals:
     Min       1Q   Median       3Q      Max
-2.88651 -0.13678 -0.00285  0.19673  0.34909

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.238e+04  4.614e+02 -26.835   <2e-16 ***
t            1.246e+01  4.671e-01  26.685   <2e-16 ***
t2          -3.135e-03  1.182e-04 -26.524   <2e-16 ***
sin.t       -1.770e-02  1.924e-02  -0.920    0.358
cos.t       -1.035e-02  1.922e-02  -0.538    0.591
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2969 on 472 degrees of freedom
Multiple R-squared:  0.9081,     Adjusted R-squared:  0.9073
F-statistic:  1166 on 4 and 472 DF,  p-value: < 2.2e-16
```

< 3. Exponential Smoothing >

```
> beer<-read.csv("beer.csv",header=T,dec=",",sep=";")
> beer<-ts(beer[,1],start=1956,freq=12)
>
> HoltWinters(beer)
Holt-Winters exponential smoothing with trend and additive seasonal
component.

Call:
HoltWinters(x = beer)

Smoothing parameters:
 alpha:  1
 beta :  0
 gamma:  0

Coefficients:
             [,1]
a    4.770000e+02
b    1.000000e+00
s1   9.621933e-16
s2  -8.141636e-16
s3  -8.141636e-16
s4   9.621933e-16
s5  -8.141636e-16
s6  -8.141636e-16
s7   9.621933e-16
s8   2.738550e-15
s9  -8.141636e-16
s10 -8.141636e-16
s11  7.401487e-17
s12 -8.141636e-16
>
> plot(beer)
> lines(HoltWinters(beer)$fitted,col="red")
Error in xy.coords(x, y) : 'x' and 'y' lengths differ
```

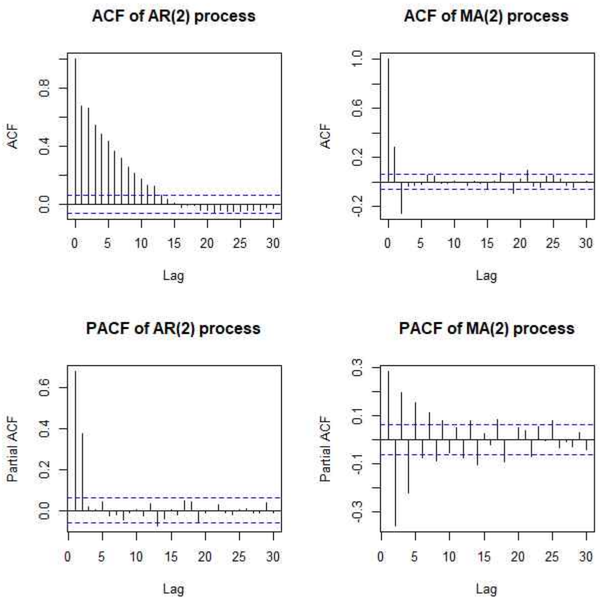



```
> beer.hw<-HoltWinters(beer)
>
> predict(beer.hw,n.ahead=12)
     Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
1995                                     478 479 480
1996 481 482 483 484 485 486 487 488 489
>
> plot(beer,xlim=c(1956,1999))
> lines(predict(beer.hw,n.ahead=48),col=2)
```

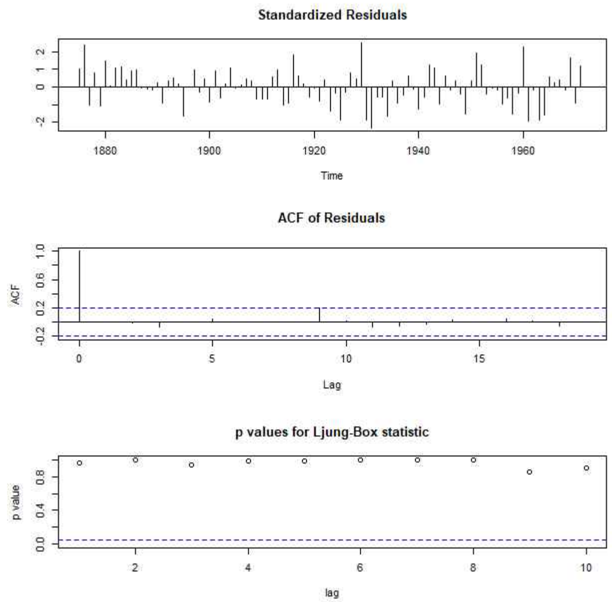


< 4. ARIMA-Models >

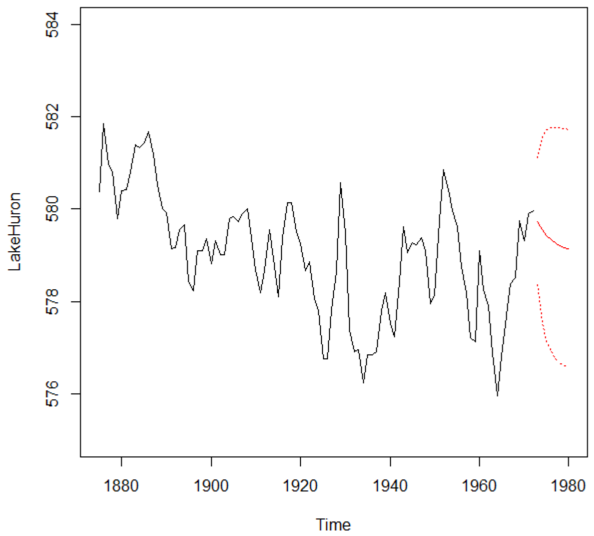
```
> sim.ar<-arima.sim(list(ar=c(0.4,0.4)),n=1000)
> sim.ma<-arima.sim(list(ma=c(0.6,-0.4)),n=1000)
> par(mfrow=c(2,2))
> acf(sim.ar,main="ACF of AR(2) process")
> acf(sim.ma,main="ACF of MA(2) process")
> pacf(sim.ar,main="PACF of AR(2) process")
> pacf(sim.ma,main="PACF of MA(2) process")
```




```
> arima(data,order=c(p,d,q))
Error in attr(data, "tsp") <- c(start, end, frequency) :
  객체가 행렬이 아닙니다
>
> data(LakeHuron)
> fit<-arima(LakeHuron,order=c(1,0,1))
>
> fit<-arima(LakeHuron,order=c(1,0,1))
> tsdiag(fit)
```




```
> Box.test(fit$residuals,lag=1)

        Box-Pierce test

data:  fit$residuals
X-squared = 0.0021379, df = 1, p-value = 0.9631

>
> fit<-arima(LakeHuron,order=c(1,0,1))
>
> LH.pred<-predict(fit,n.ahead=8)
>
> plot(LakeHuron,xlim=c(1875,1980),ylim=c(575,584))
> LH.pred<-predict(fit,n.ahead=8)
> lines(LH.pred$pred,col="red")
> lines(LH.pred$pred+2*LH.pred$se,col="red",lty=3)
> lines(LH.pred$pred-2*LH.pred$se,col="red",lty=3)
```



584
582
580
578
576
LakeHuron
1880
1900
1920
1940
1960
1980
Time