STAT4870 Chapter 3 (2)

Anton Yang

2024-09-19

3.2 Explanatory Data Analysis

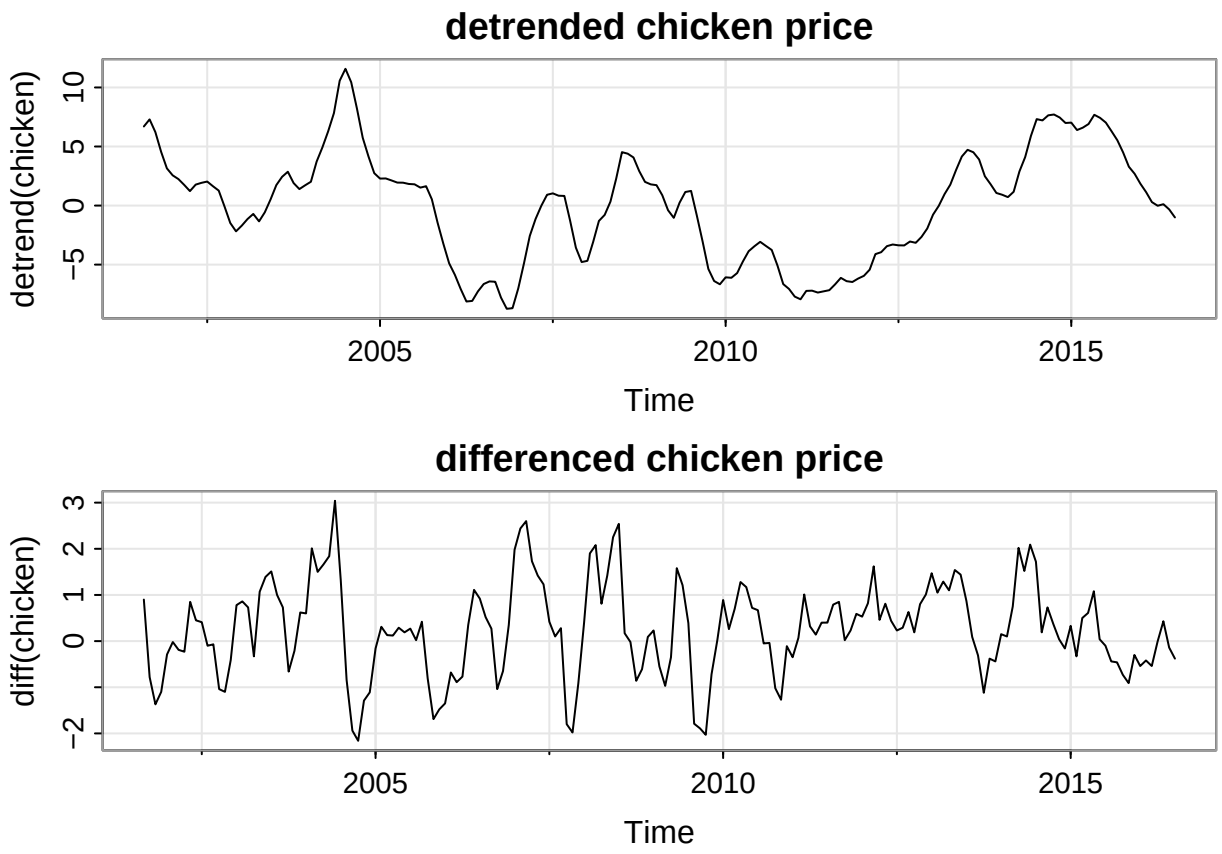
```
library(astsa)
library(xts)
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

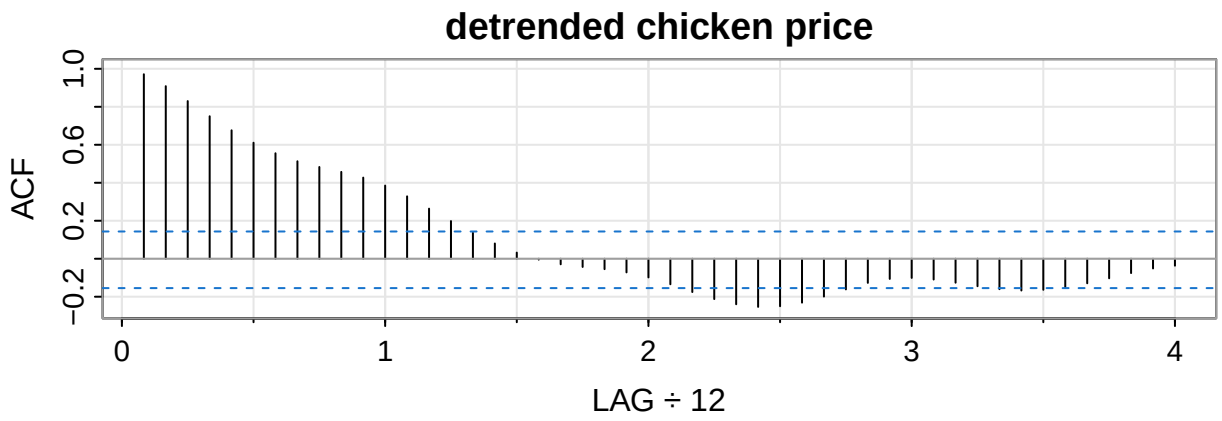
```
data(chicken)
par(mfrow=2:1) # plot transformed data
tsplot(detrend(chicken), main="detrended chicken price" )
tsplot(diff(chicken), main="differenced chicken price" )
```

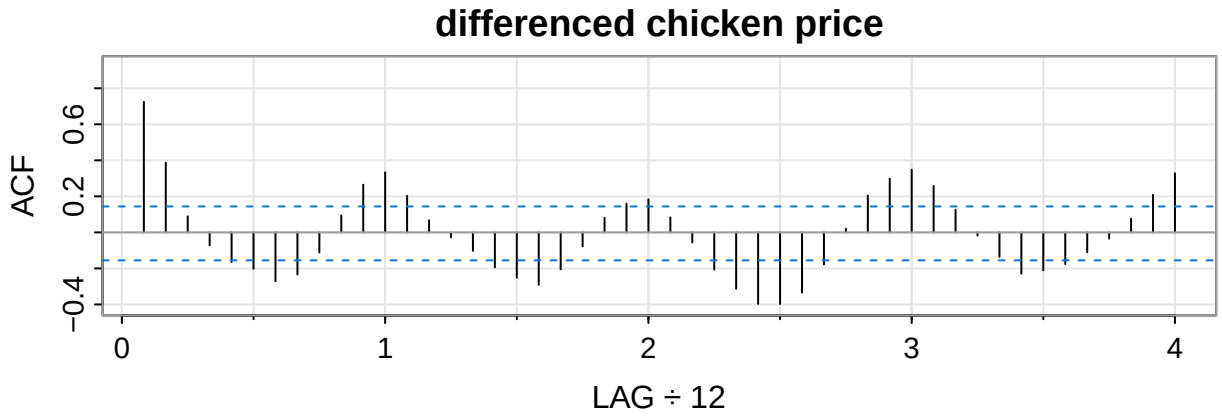



```
acf1(detrend(chicken), main="detrended chicken price")
```

```
##  [1]  0.97  0.91  0.83  0.75  0.68  0.61  0.56  0.51  0.48  0.46  0.43  0.39
## [13]  0.33  0.26  0.20  0.14  0.08  0.03  0.00 -0.03 -0.04 -0.05 -0.07 -0.10
## [25] -0.13 -0.18 -0.21 -0.24 -0.25 -0.25 -0.23 -0.20 -0.16 -0.13 -0.11 -0.10
## [37] -0.11 -0.13 -0.14 -0.16 -0.17 -0.16 -0.15 -0.13 -0.10 -0.08 -0.05 -0.04
```

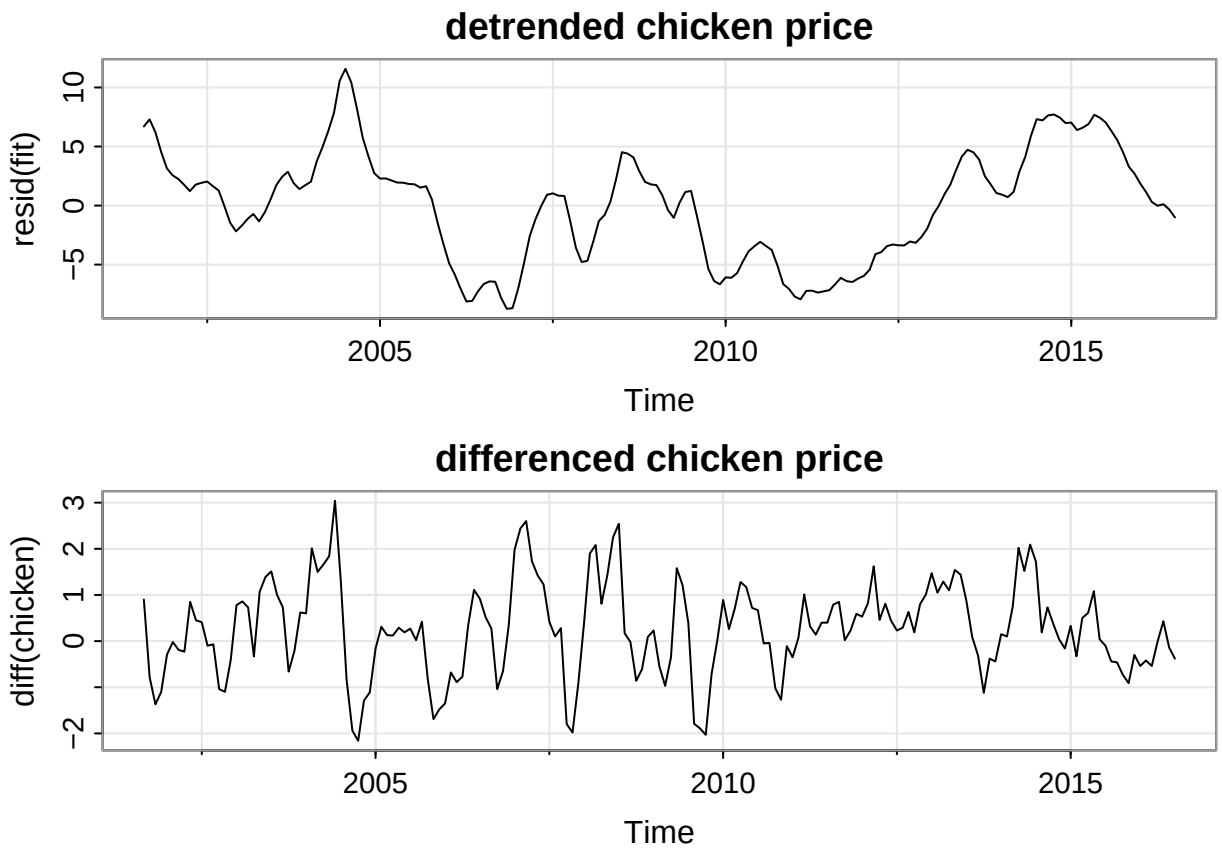
```
acf1(diff(chicken), main="differenced chicken price")
```






```
##  [1]  0.72  0.39  0.09 -0.07 -0.16 -0.20 -0.27 -0.23 -0.11  0.09  0.26  0.33
## [13]  0.20  0.07 -0.03 -0.10 -0.19 -0.25 -0.29 -0.20 -0.08  0.08  0.16  0.18
## [25]  0.08 -0.06 -0.21 -0.31 -0.40 -0.40 -0.33 -0.18  0.02  0.20  0.30  0.35
## [37]  0.26  0.13 -0.02 -0.14 -0.23 -0.21 -0.18 -0.11 -0.03  0.08  0.21  0.33
```

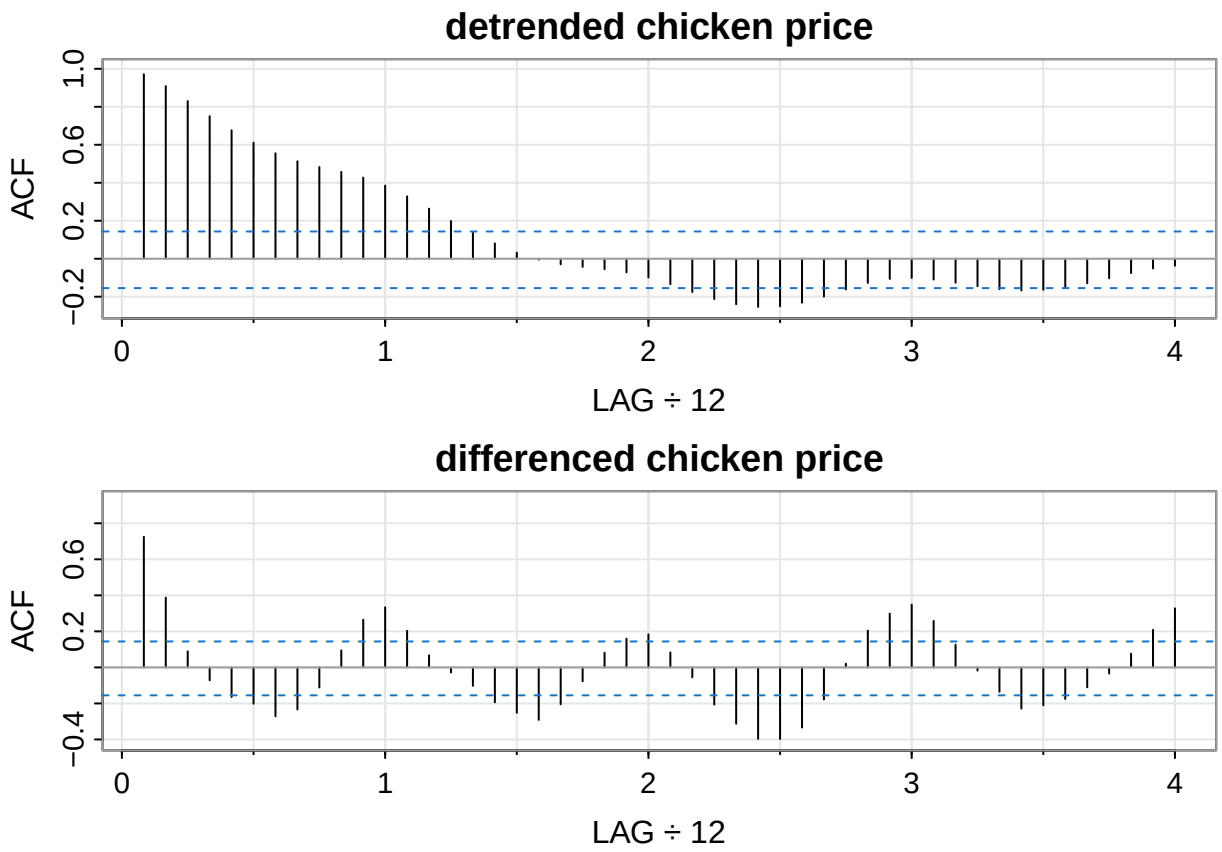
```
fit <- lm(chicken~time(chicken), na.action=NULL)
tsplot(resid(fit), main="detrended chicken price")
tsplot(diff(chicken), main="differenced chicken price")
```




```
acf1(resid(fit), 48, main="detrended chicken price")
```

```
##  [1]  0.97  0.91  0.83  0.75  0.68  0.61  0.56  0.51  0.48  0.46  0.43  0.39
## [13]  0.33  0.26  0.20  0.14  0.08  0.03  0.00 -0.03 -0.04 -0.05 -0.07 -0.10
## [25] -0.13 -0.18 -0.21 -0.24 -0.25 -0.25 -0.23 -0.20 -0.16 -0.13 -0.11 -0.10
## [37] -0.11 -0.13 -0.14 -0.16 -0.17 -0.16 -0.15 -0.13 -0.10 -0.08 -0.05 -0.04
```

```
acf1(diff(chicken), 48, main="differenced chicken price")
```

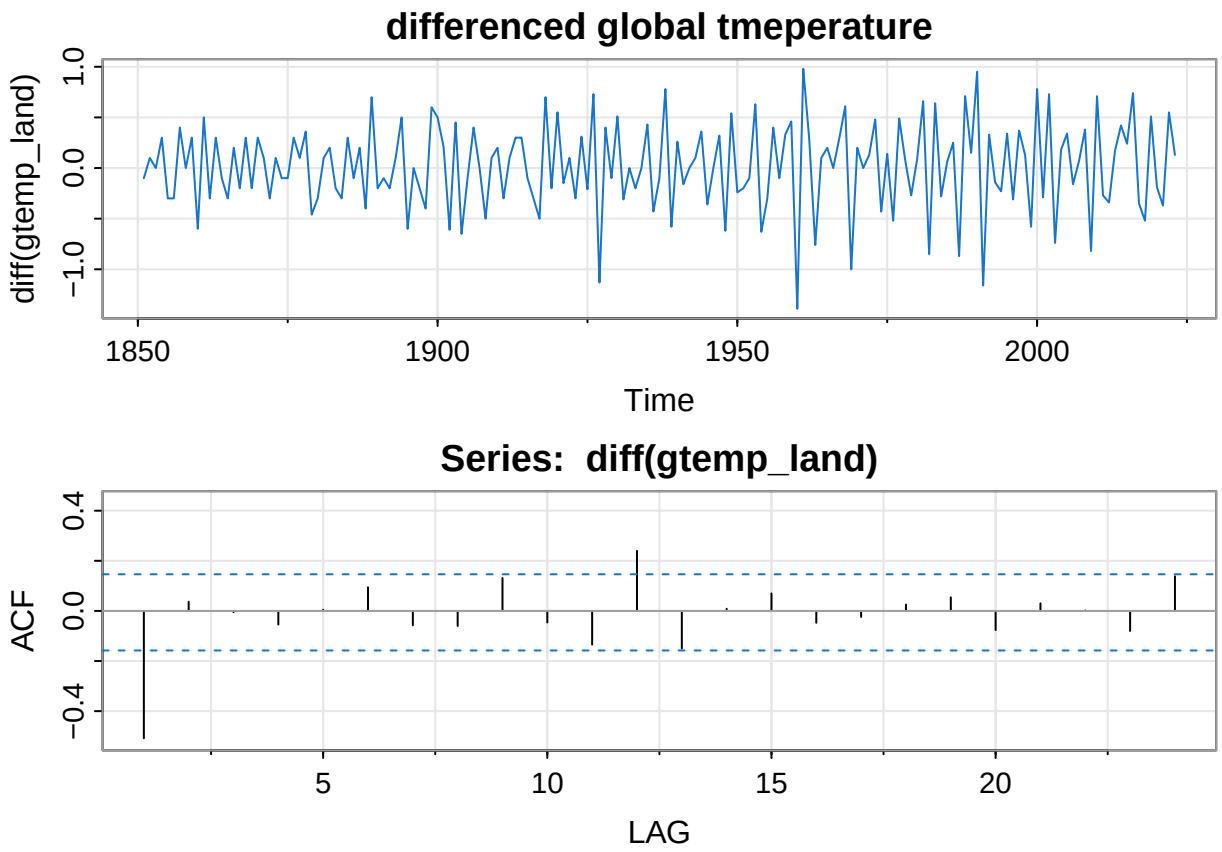



```
##  [1]  0.72  0.39  0.09 -0.07 -0.16 -0.20 -0.27 -0.23 -0.11  0.09  0.26  0.33
## [13]  0.20  0.07 -0.03 -0.10 -0.19 -0.25 -0.29 -0.20 -0.08  0.08  0.16  0.18
## [25]  0.08 -0.06 -0.21 -0.31 -0.40 -0.40 -0.33 -0.18  0.02  0.20  0.30  0.35
## [37]  0.26  0.13 -0.02 -0.14 -0.23 -0.21 -0.18 -0.11 -0.03  0.08  0.21  0.33
```

```
par(mfrow=c(2,1))
tsplot(diff(gtemp_land), col=4, main="differenced global tmeperature")
mean(diff(gtemp_land))
```

```
## [1] 0.01595376
```

```
acf1(diff(gtemp_land))
```



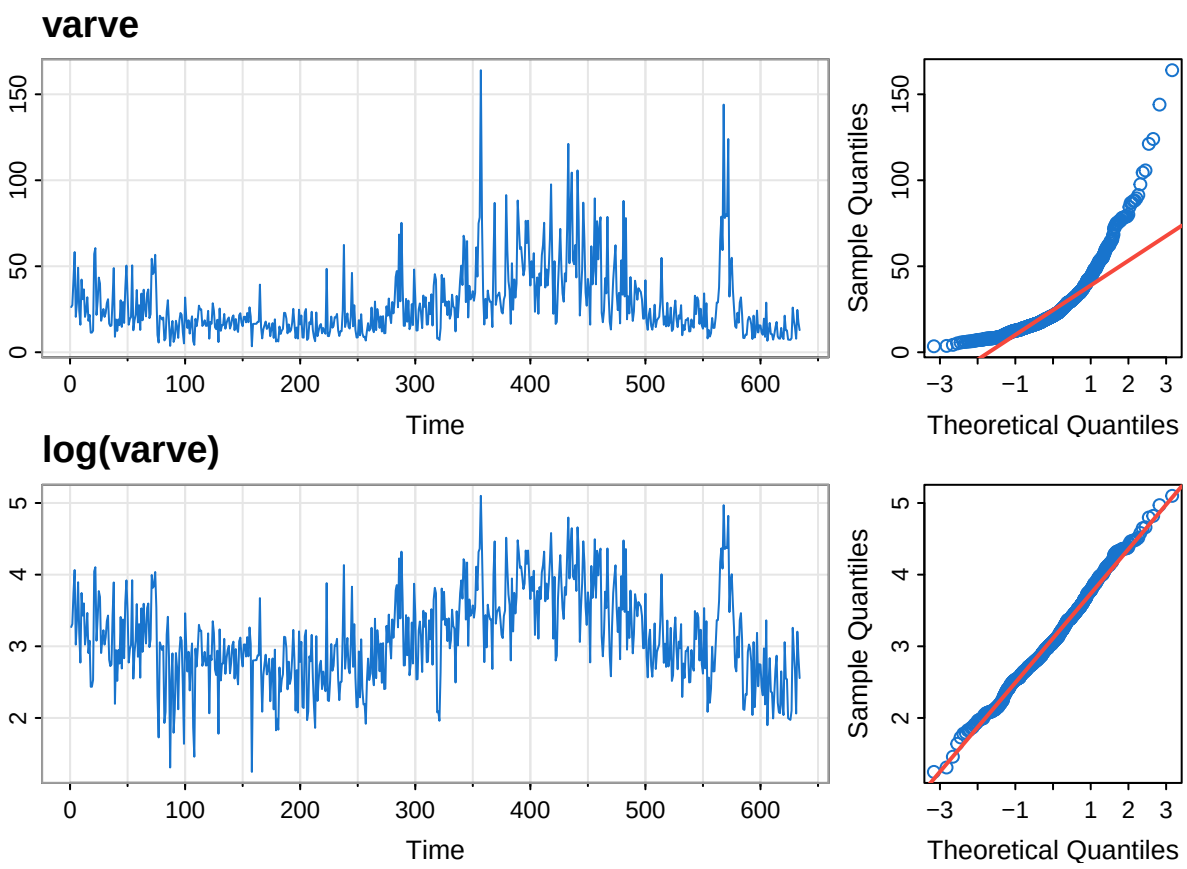

```
##  [1] -0.51  0.04  0.00 -0.05  0.00  0.09 -0.06 -0.06  0.13 -0.05 -0.13  0.24
## [13] -0.15  0.01  0.07 -0.05 -0.02  0.03  0.05 -0.08  0.03  0.00 -0.08  0.14
```

```
mean(window(diff(gtemp_land), start=1980))
```

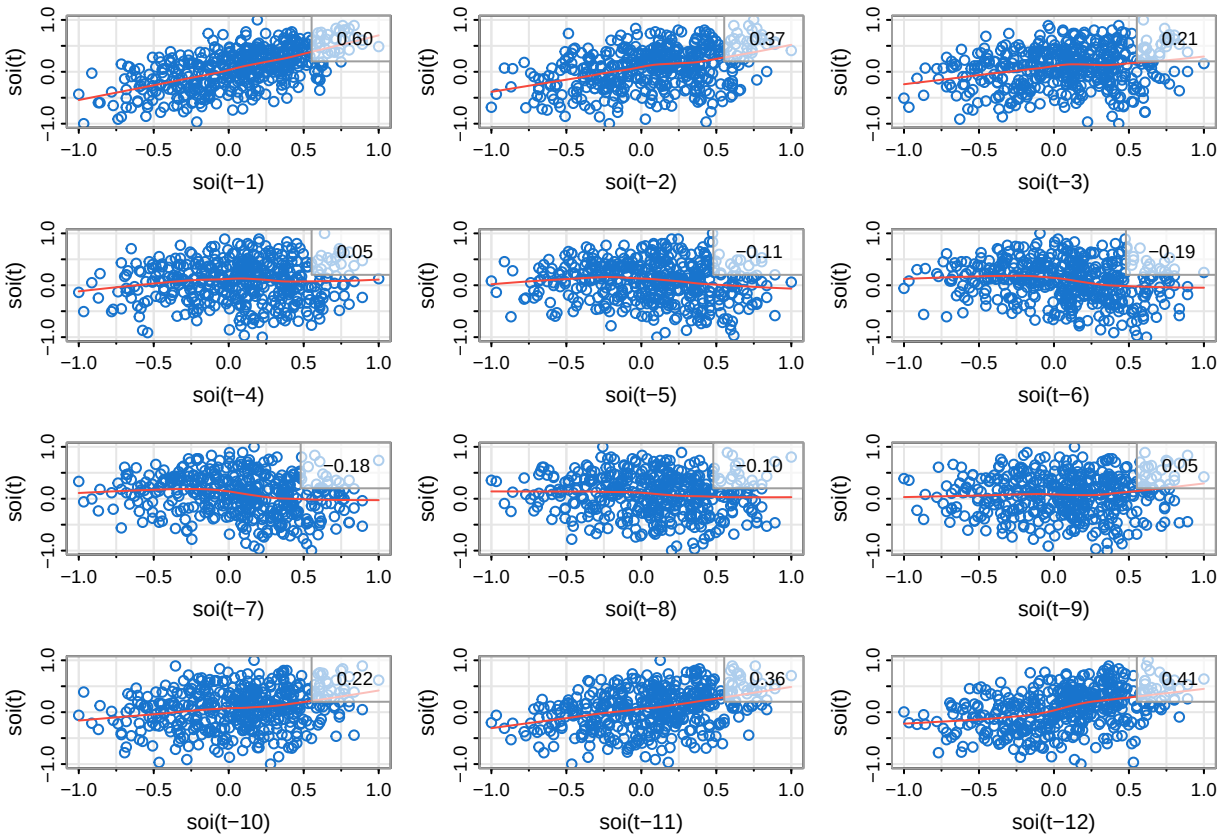
```
## [1] 0.04909091
```

```
par(mfrow=c(1,1))
```

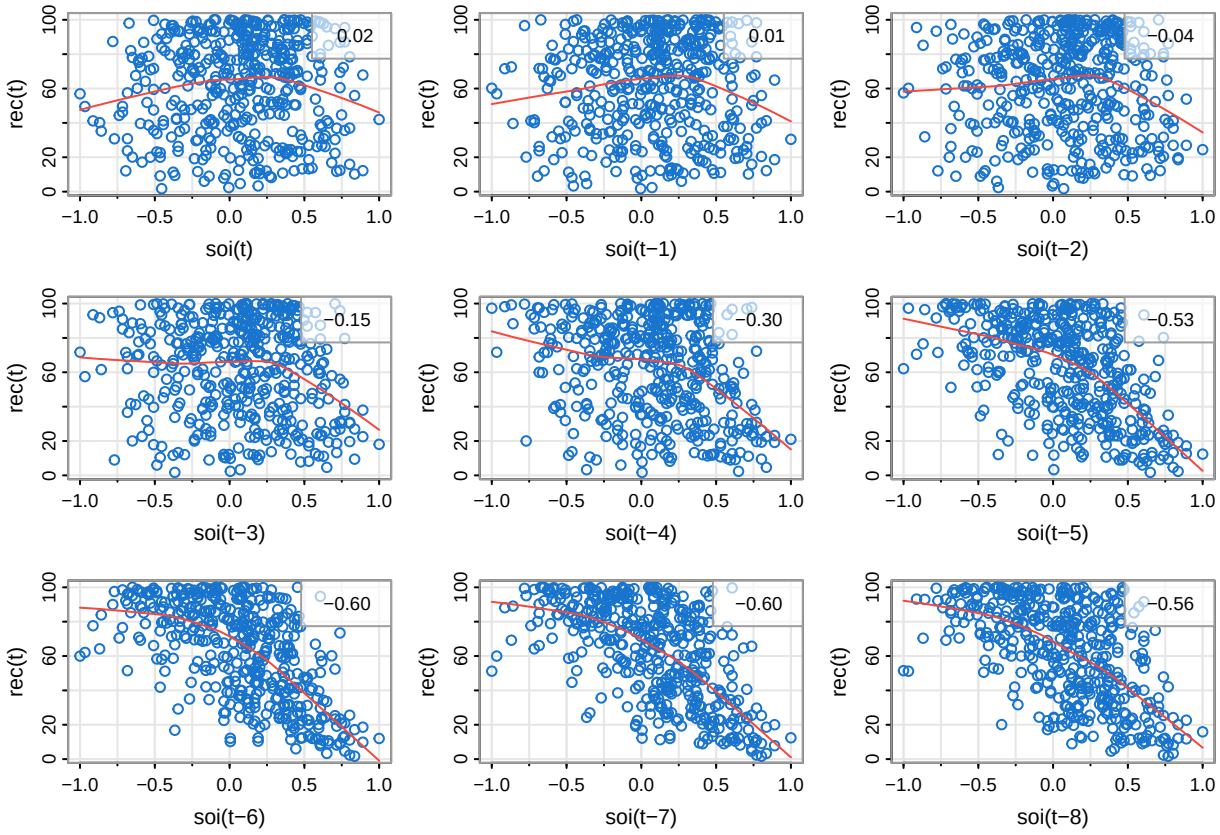
```
layout(matrix(1:4,2), widths=c(2.5,1))
par(oma=rep(.2, 4))
tsplot(varve, main="", ylab="", col=4, margin=0)
mtext("varve", side=3, line=.5, cex=1.2, font=2, adj=0)
tsplot(log(varve), main="", ylab="", col=4, margin=0)
mtext("log(varve)", side=3, line=.5, cex=1.2, font=2, adj=0)
qqnorm(varve, main="", col=4); qqline(varve, col=2, lwd=2)
qqnorm(log(varve), main="", col=4); qqline(log(varve), col=2, lwd=2)
```




```
lag1.plot(soi, 12, col=4, cex=1)
```



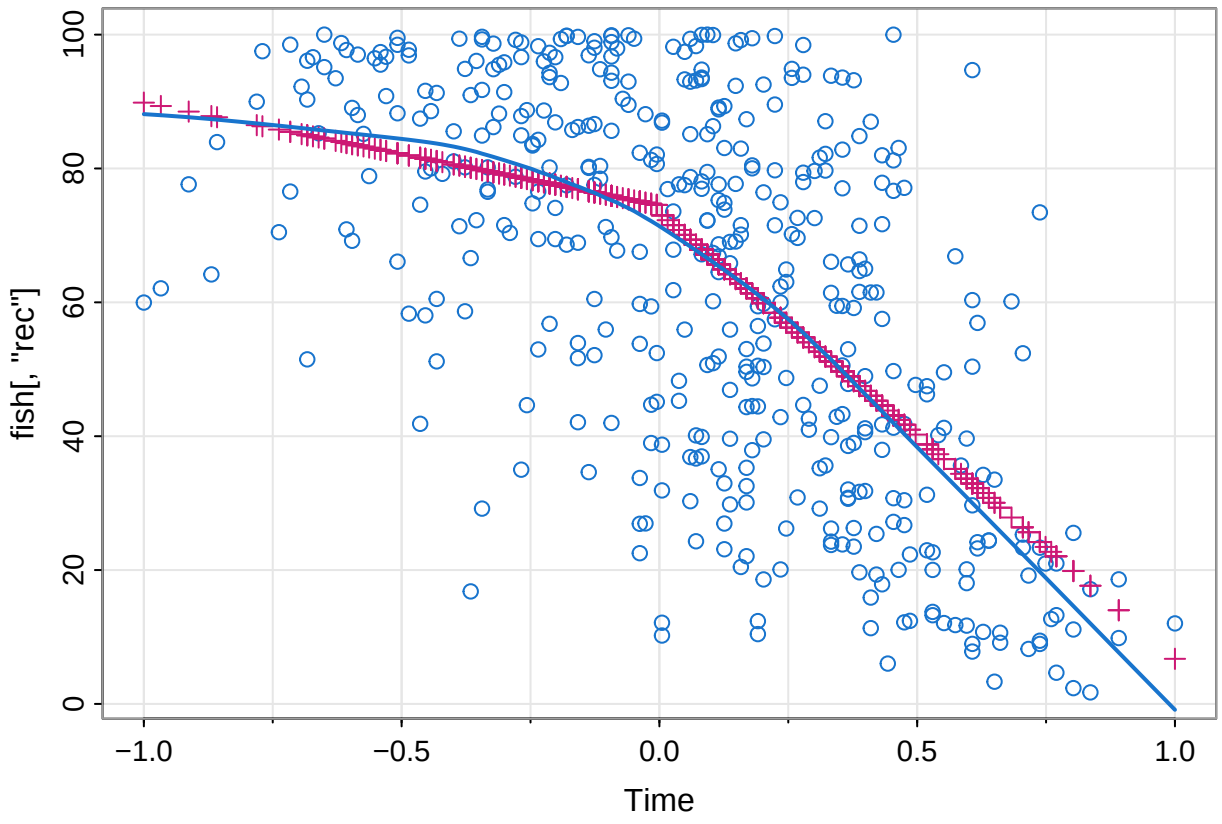

```
lag2.plot(soi, rec, 8, col=4, cex=1)
```



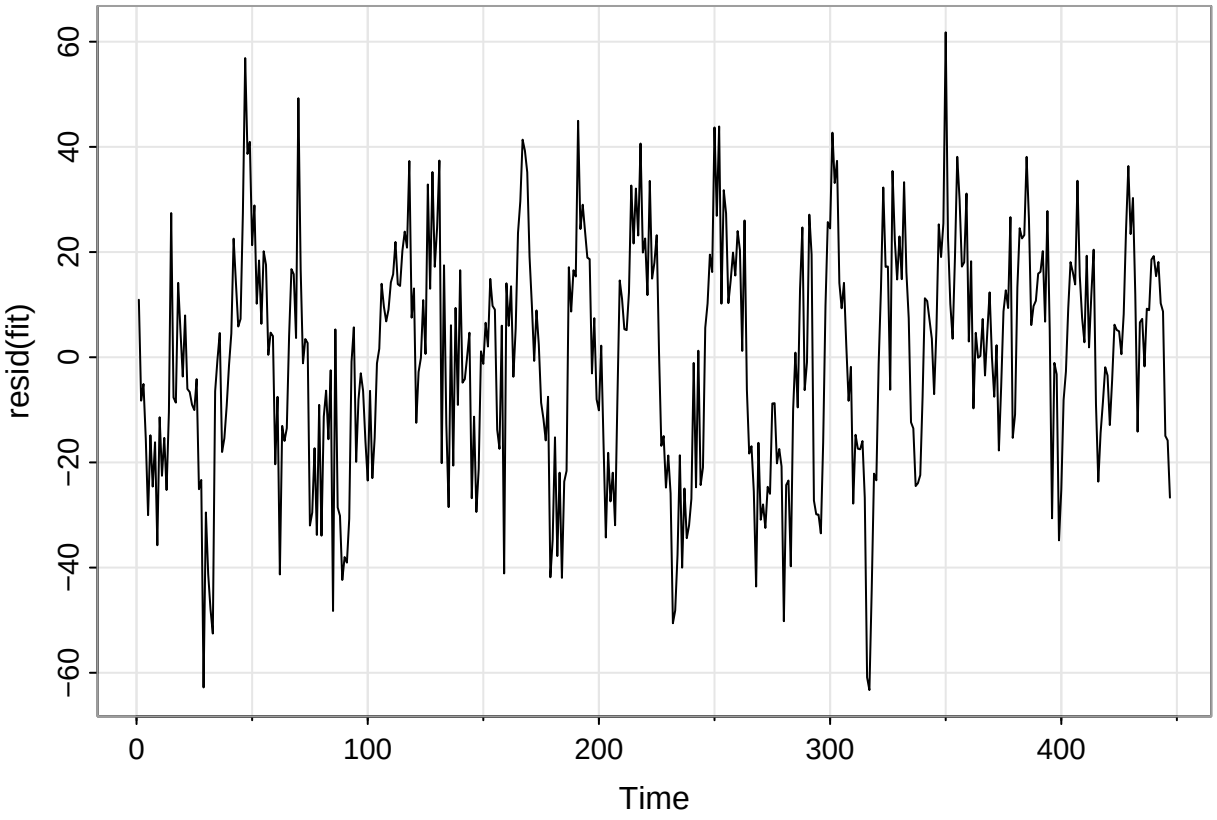

```
dummy <- ifelse(soi<0, 0, 1)
fish <- ts.intersect(rec, soiL6=lag(soi,-6), dL6=lag(dummy,-6))
summary(fit <- lm(rec~ soiL6*dL6, data=fish, na.action=NULL))
```

```
##
## Call:
## lm(formula = rec ~ soiL6 * dL6, data = fish, na.action = NULL)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -63.291 -15.821   2.224  15.791  61.788
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   74.479      2.865  25.998  < 2e-16 ***
## soiL6        -15.358      7.401  -2.075   0.0386 *
## dL6           -1.139      3.711  -0.307   0.7590
## soiL6:dL6    -51.244      9.523  -5.381  1.2e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 21.84 on 443 degrees of freedom
## Multiple R-squared:  0.4024, Adjusted R-squared:  0.3984
## F-statistic: 99.43 on 3 and 443 DF,  p-value: < 2.2e-16
```

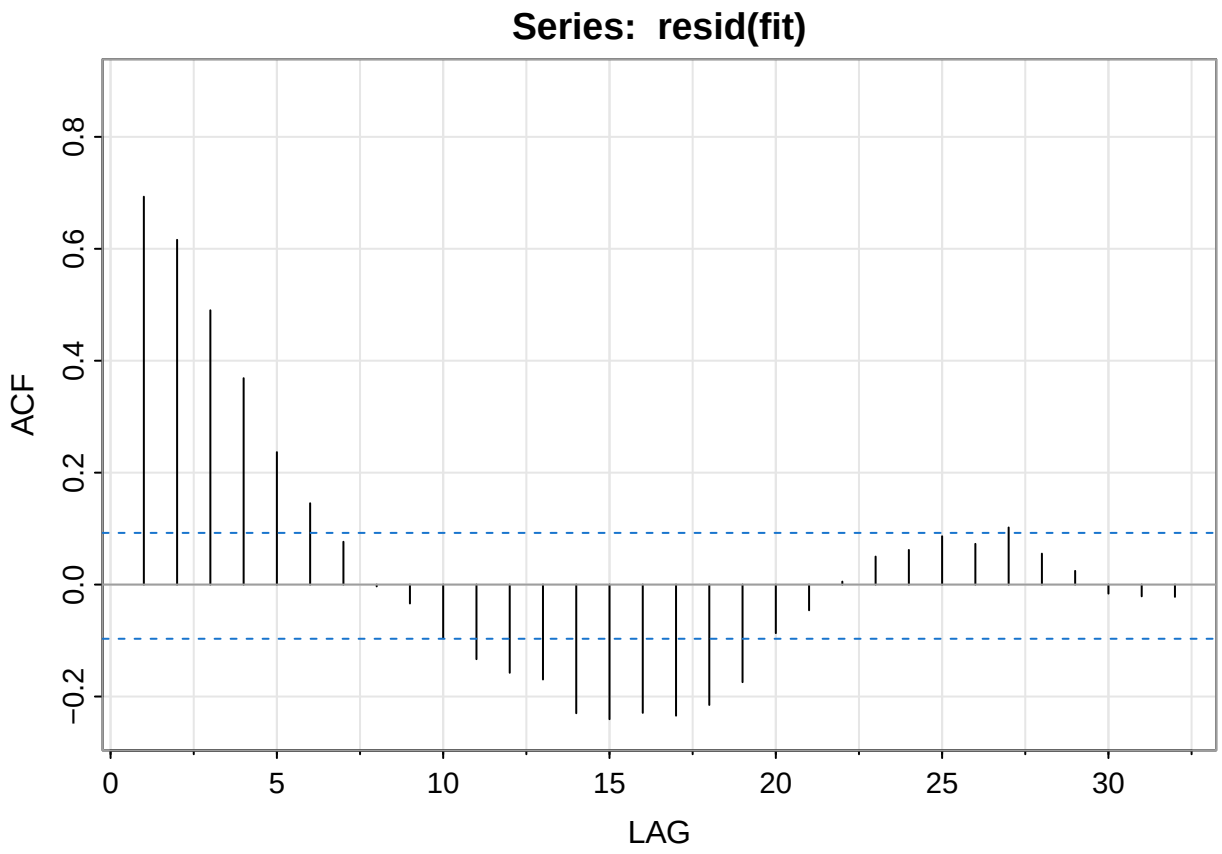
```
tsplot(fish[,'soiL6'], fish[,'rec'], type='p', col=4)
points(fish[,'soiL6'], fitted(fit), pch=3, col=6)
lines(lowess(fish[,'soiL6'], fish[,'rec']), col=4, lwd=2)
```




```
tsplot(resid(fit))
```




```
acf1(resid(fit))
```



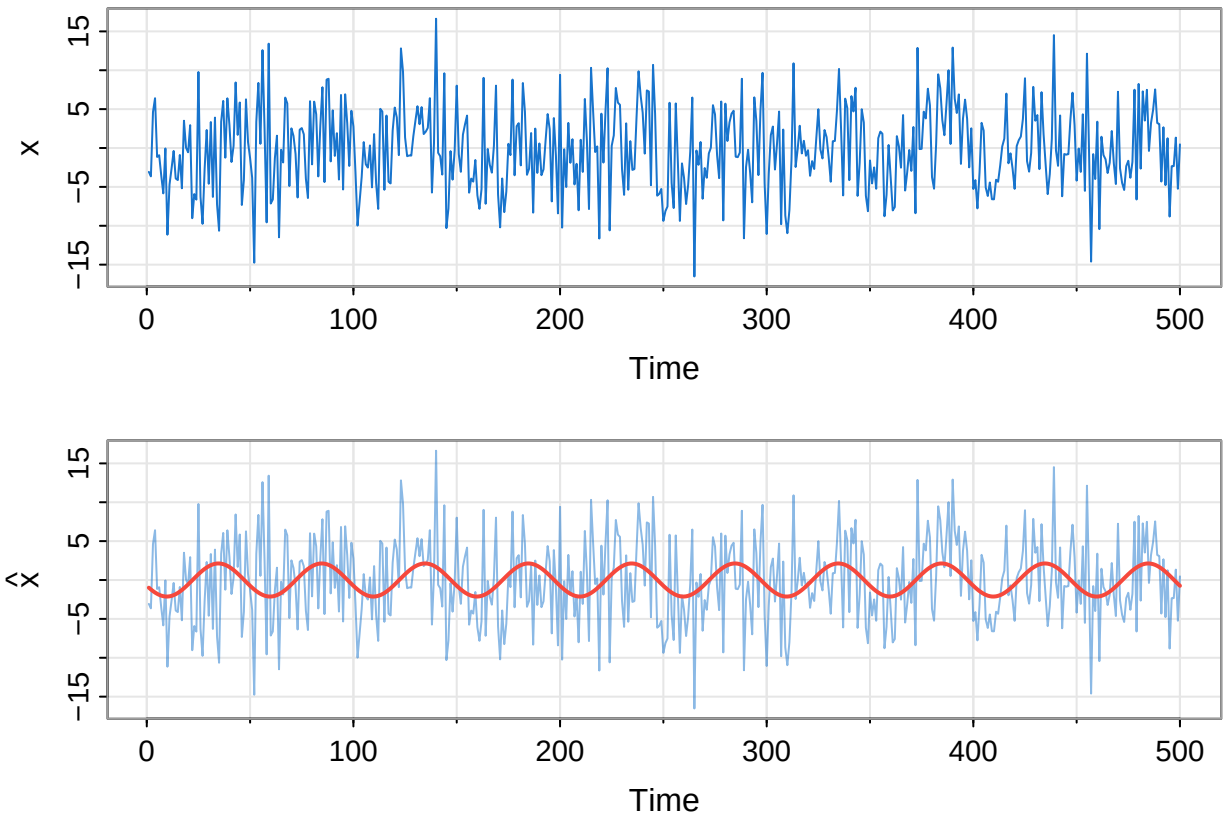

```
##  [1]  0.69  0.62  0.49  0.37  0.24  0.15  0.08  0.00 -0.03 -0.10 -0.13 -0.16
## [13] -0.17 -0.23 -0.24 -0.23 -0.23 -0.22 -0.17 -0.09 -0.05  0.01  0.05  0.06
## [25]  0.09  0.07  0.10  0.06  0.02 -0.02 -0.02 -0.02
```

```
set.seed(90210)
x <- 2*cos(2*pi*1:500/50 + .6*pi) + rnorm(500,0,5)
z1 <- cos(2*pi*1:500/50)
z2 <- sin(2*pi*1:500/50)
summary(fit <- lm(x~ 0 + z1 + z2))
```

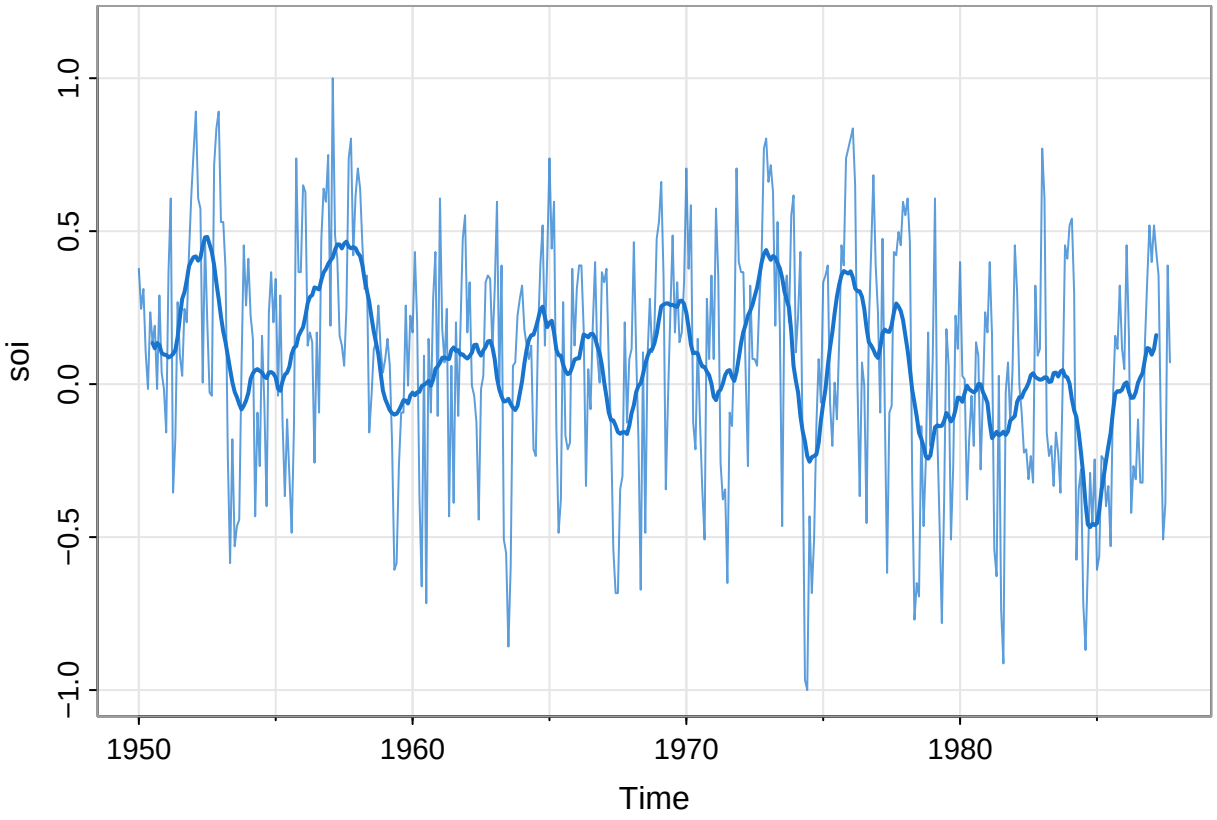
```
##
## Call:
## lm(formula = x ~ 0 + z1 + z2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.8584  -3.8525  -0.3186   3.3487  15.5440
##
## Coefficients:
##    Estimate Std. Error t value Pr(>|t|)
## z1  -0.7442     0.3274  -2.273   0.0235 *
## z2  -1.9949     0.3274  -6.093 2.23e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 5.177 on 498 degrees of freedom
## Multiple R-squared:  0.07827,    Adjusted R-squared:  0.07456
## F-statistic: 21.14 on 2 and 498 DF,  p-value: 1.538e-09
```

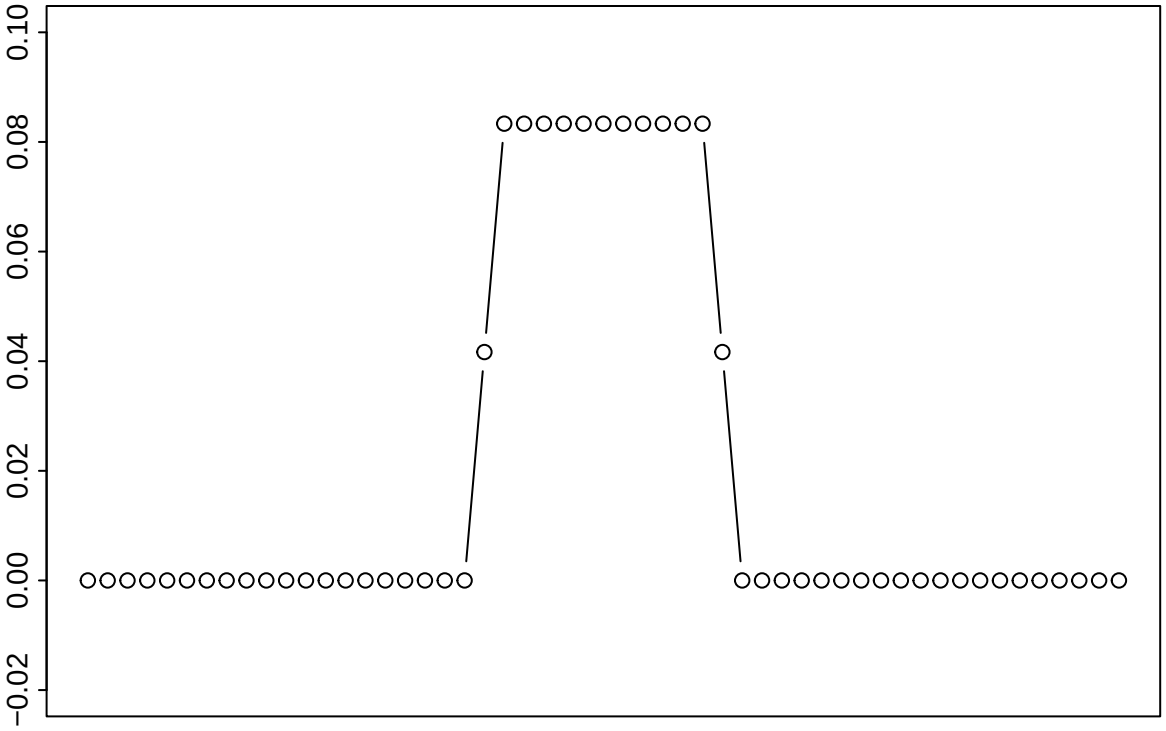
```
par(mfrow=c(2,1))
tsplot(x, col=4)
tsplot(x, ylab=expression(hat(x)), col=astsa.col(4, .5))
lines(fitted(fit), col=2, lwd=2)
```



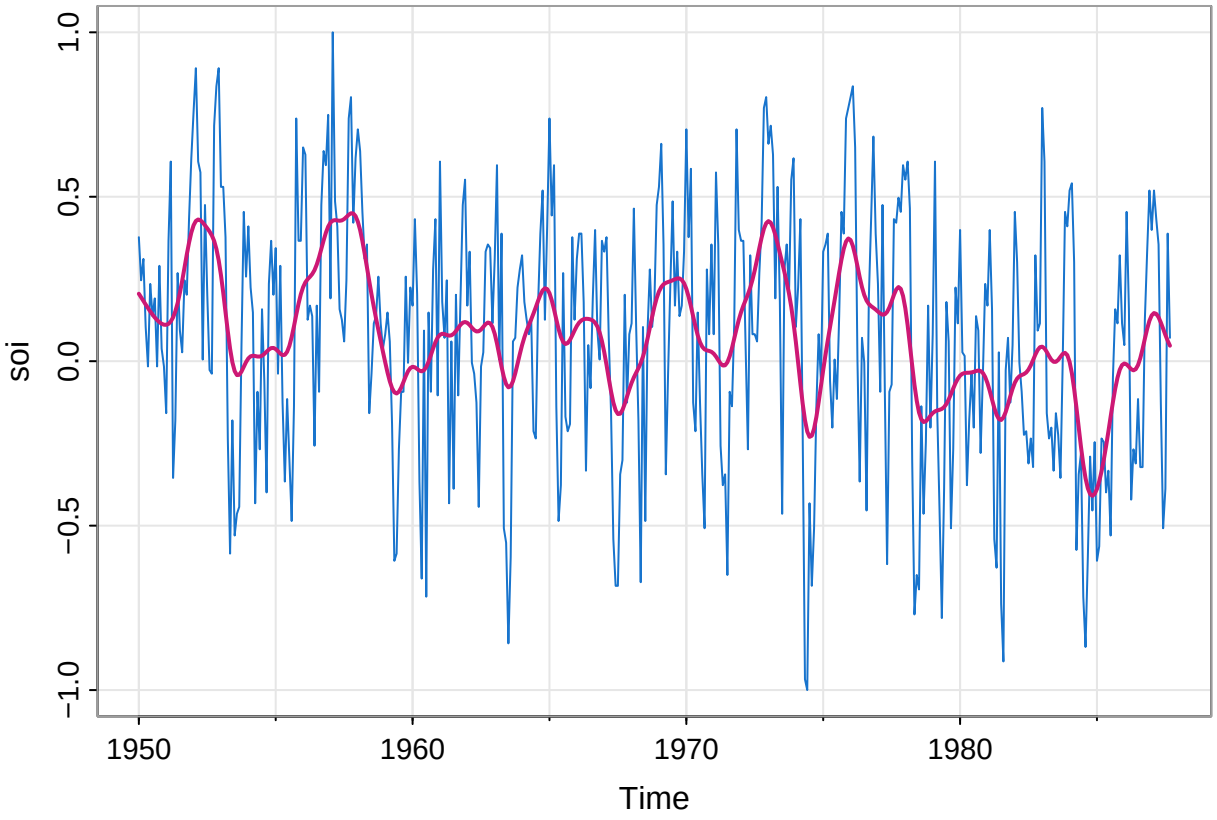

```
library(astsa)
w = c(.5, rep(1,11), .5)/12
soif = filter(soi, sides=2, filter=w)
tsplot(soi, col=astsa.col(4,.7), ylim=c(-1, 1.15))
lines(soif, lwd=2, col=4)
```



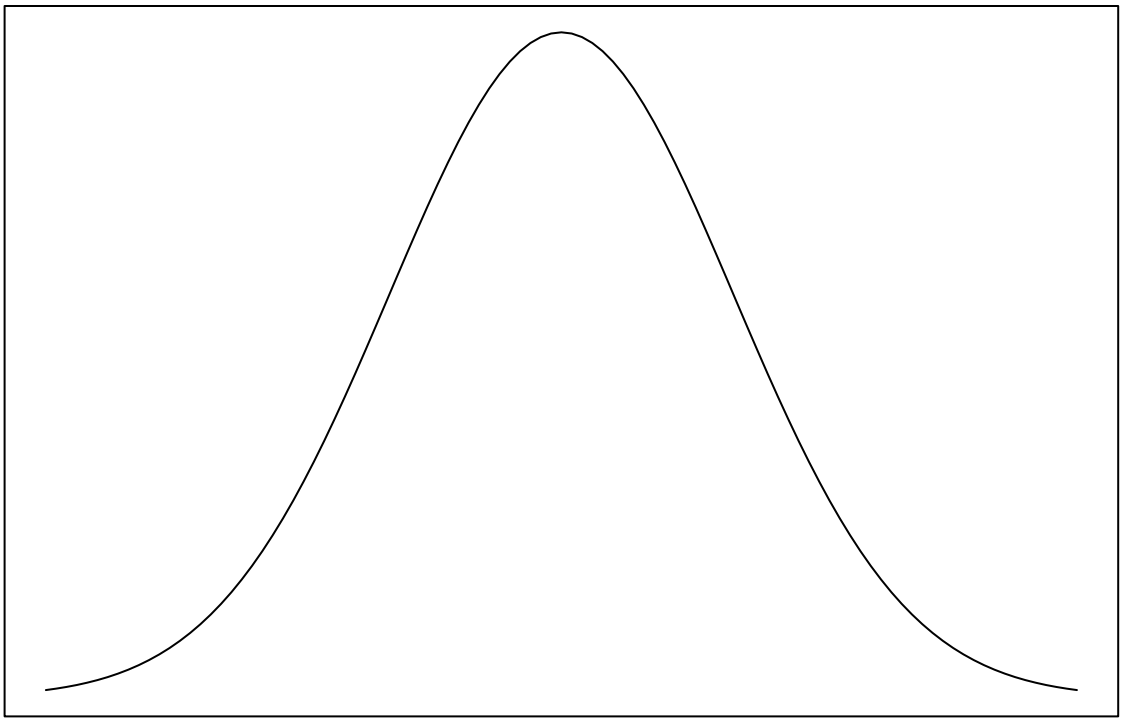

```
w1 = c(rep(0,20), w, rep(0,20))
plot(w1, type="b", ylim = c(-.02,.1), xaxt="n", ann=FALSE)
```



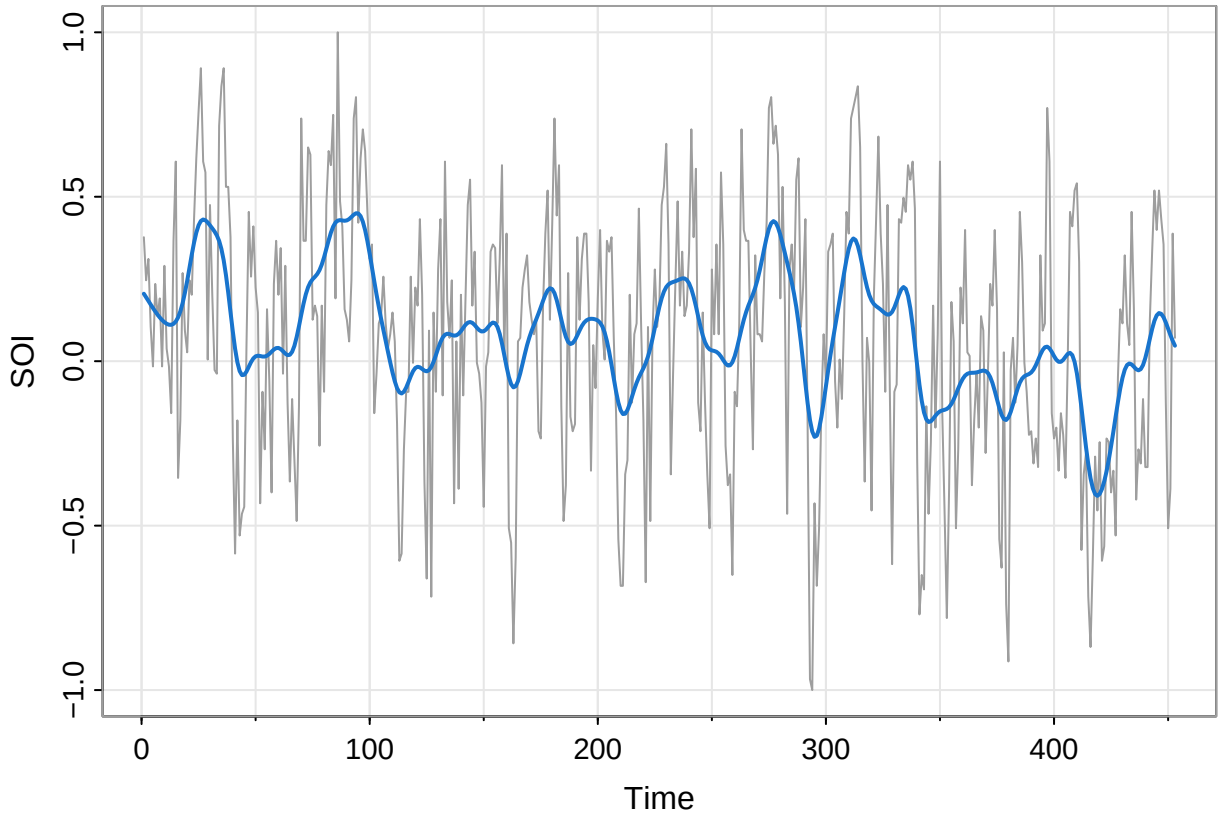

```
tsplot(soi, col=4)
lines(ksmooth(time(soi), soi, "normal", bandwidth=1), lwd=2, col=6)
```



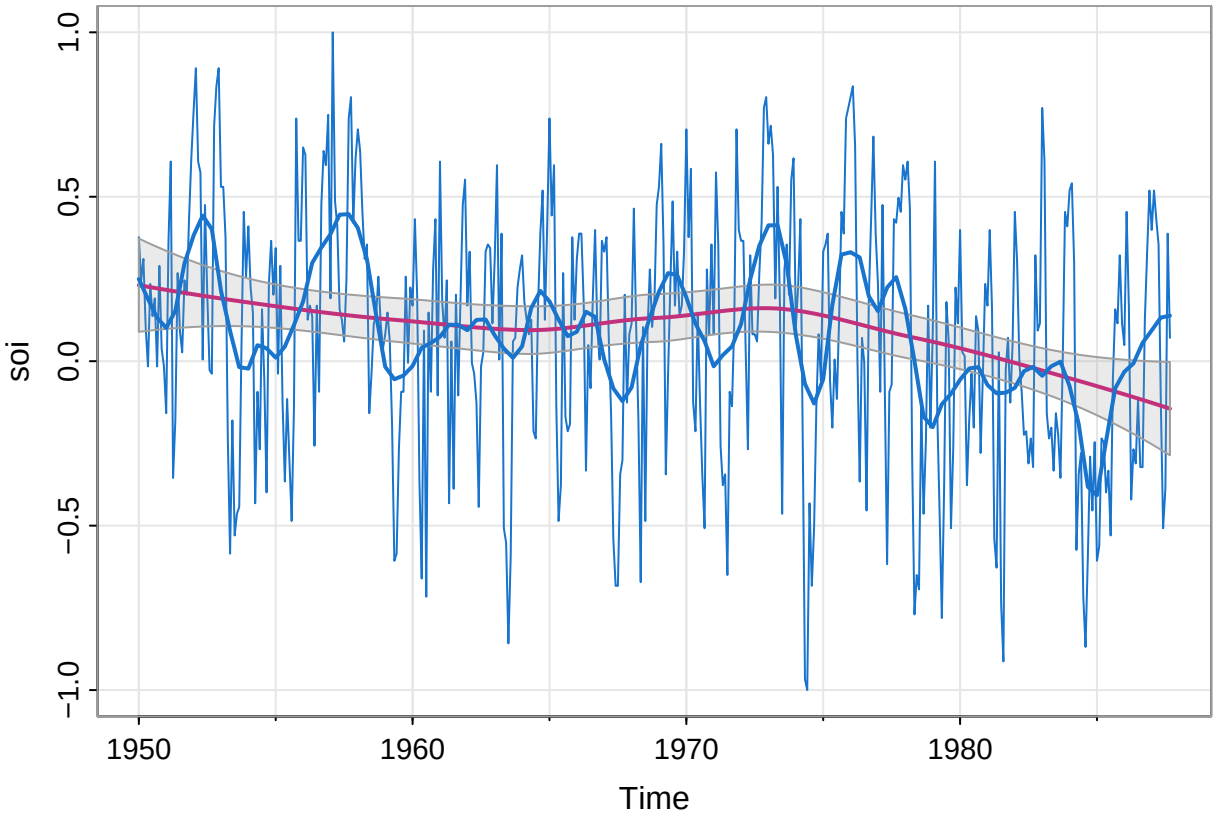

```
curve(dnorm, -3, 3, xaxt='n', yaxt='n', ann=FALSE)
```



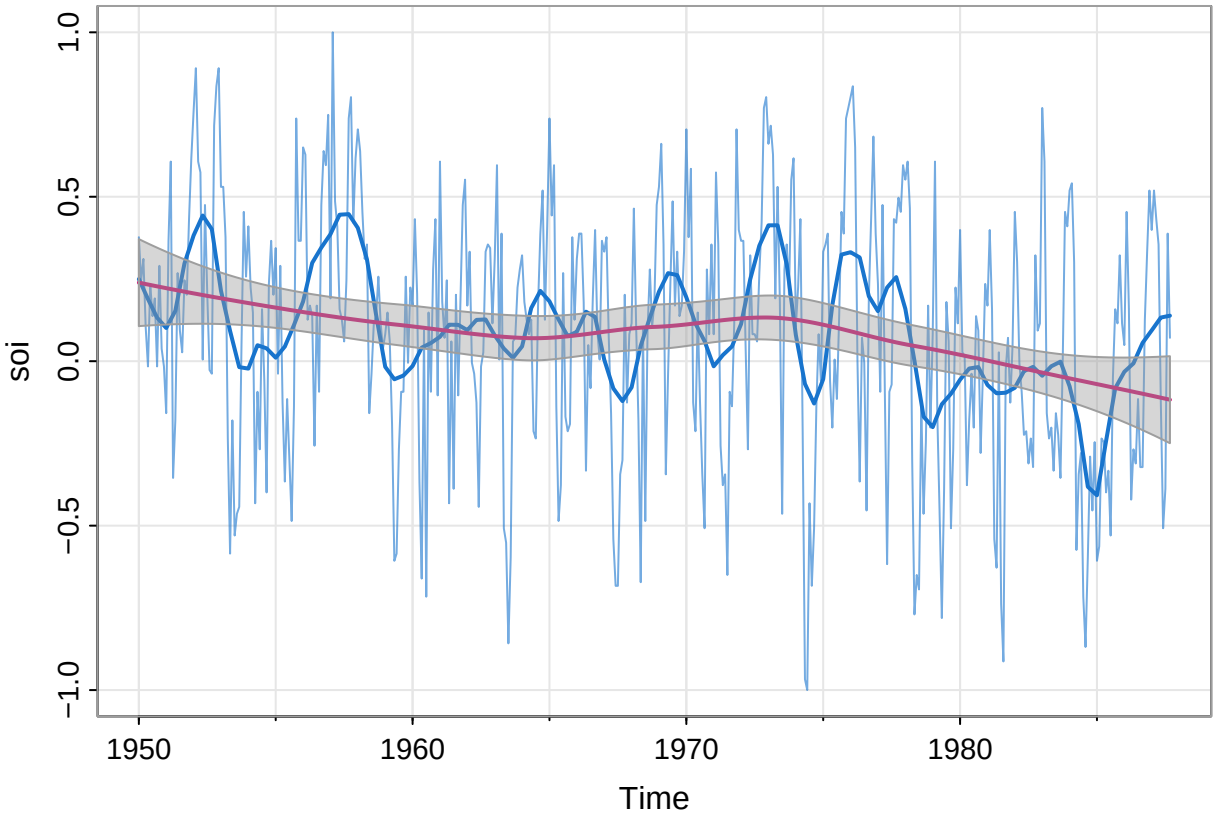
```
SOI <- ts(soi, freq=1)
tsplot(SOI, col=8) # the time scale matters (not shown)
lines(ksmooth(time(SOI), SOI, "normal", bandwidth=12), lwd=2, col=4)
```



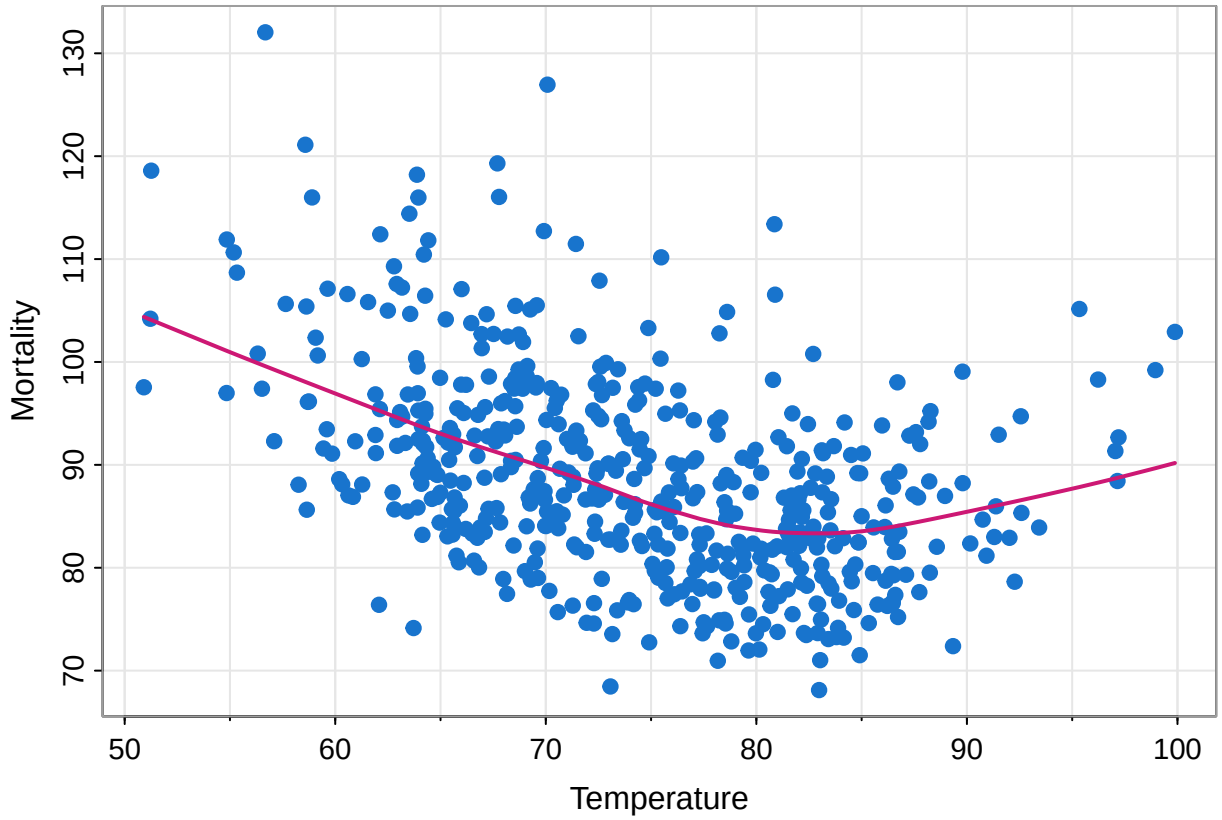

```
trend(soi, lowess=TRUE)
lines(lowess(soi, f=.05), lwd=2, col=4)
```



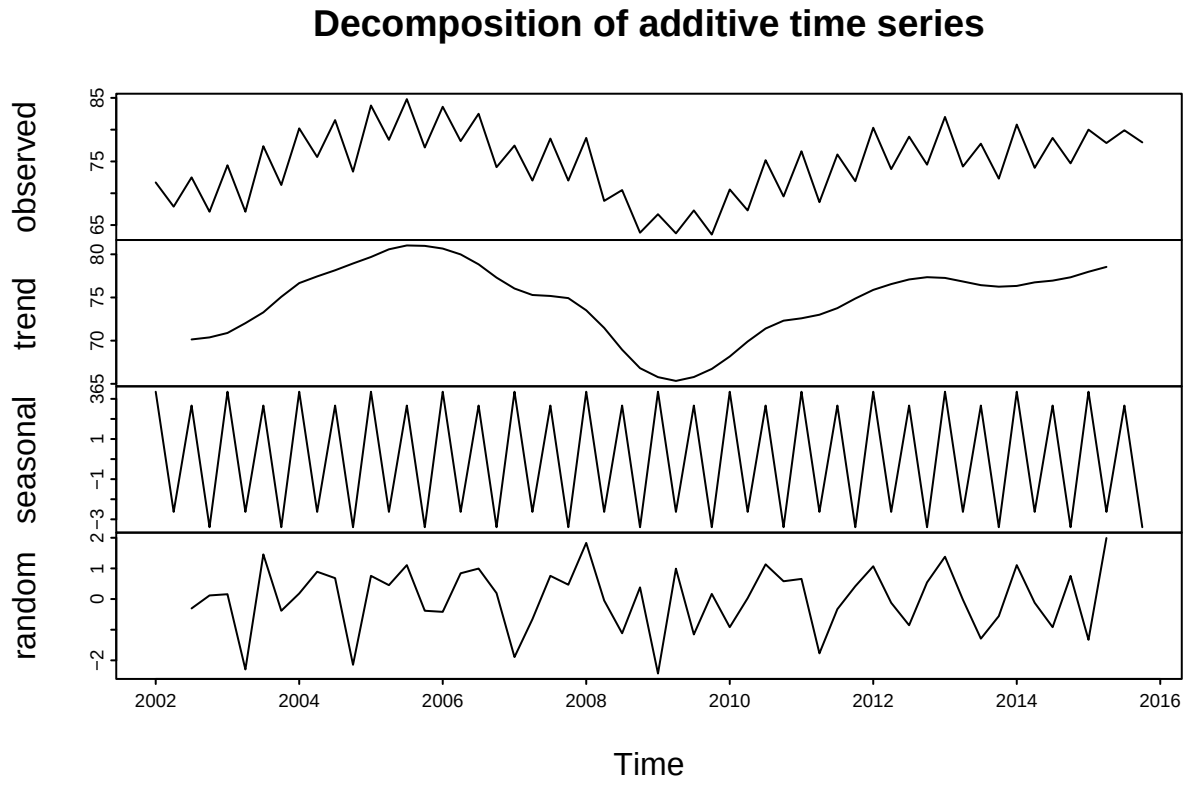

```
tsplot(soi, col=astsa.col(4,.6))
lines(lowess(soi, f=.05), lwd=2, col=4)
lo <- predict(loess(soi ~ time(soi)), se=TRUE)
trnd <- ts(lo$fit, start=1950, freq=12)
lines(trnd, col=6, lwd=2)
L <- trnd - qt(0.975, lo$df)*lo$se
U <- trnd + qt(0.975, lo$df)*lo$se
xx <- c(time(soi), rev(time(soi)))
yy <- c(L, rev(U))
polygon(xx, yy, border=8, col=gray(.6, alpha=.4) )
```



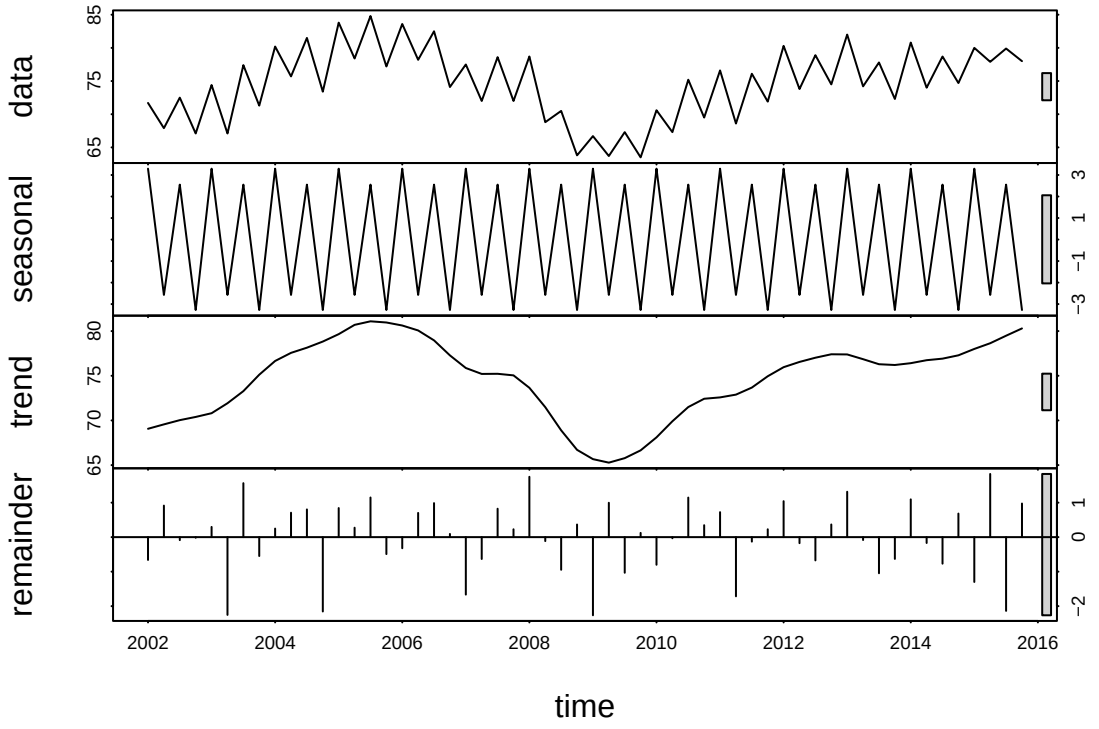

```
tsplot(tempr, cmort, type='p', pch=19, xlab="Temperature", ylab="Mortality", col=4)
lines(lowess(tempr,cmort), col=6, lwd=2)
```



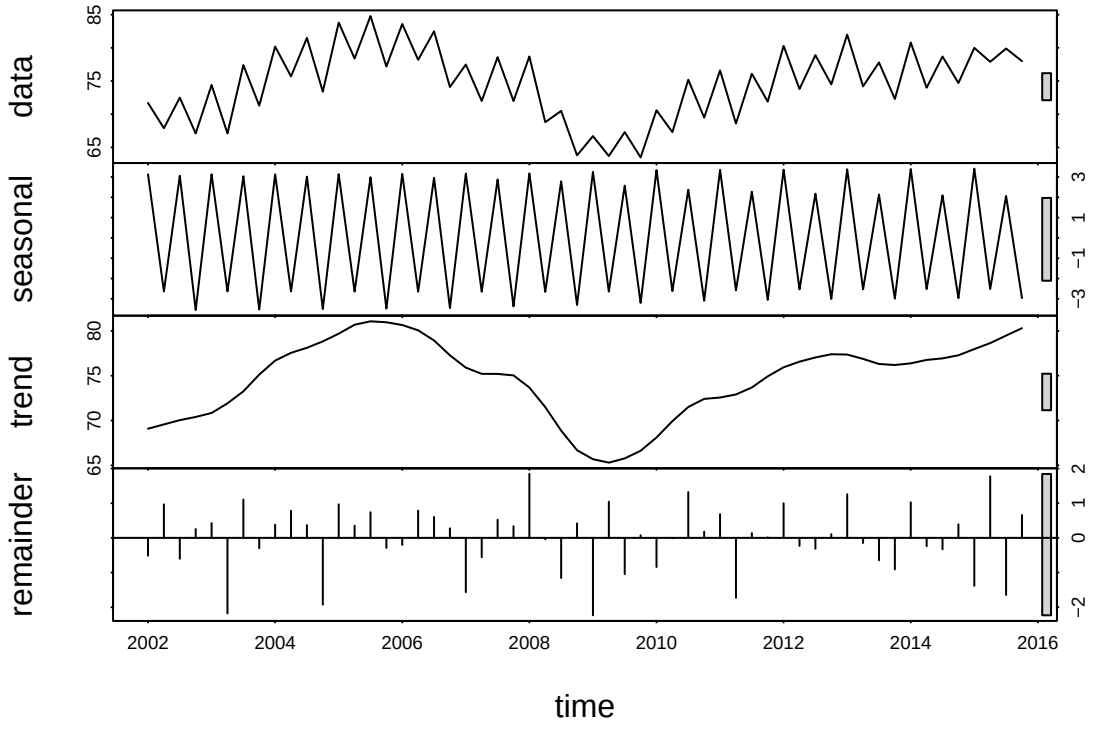

```
x = window(hor, start=2002)
plot(decompose(x))
```



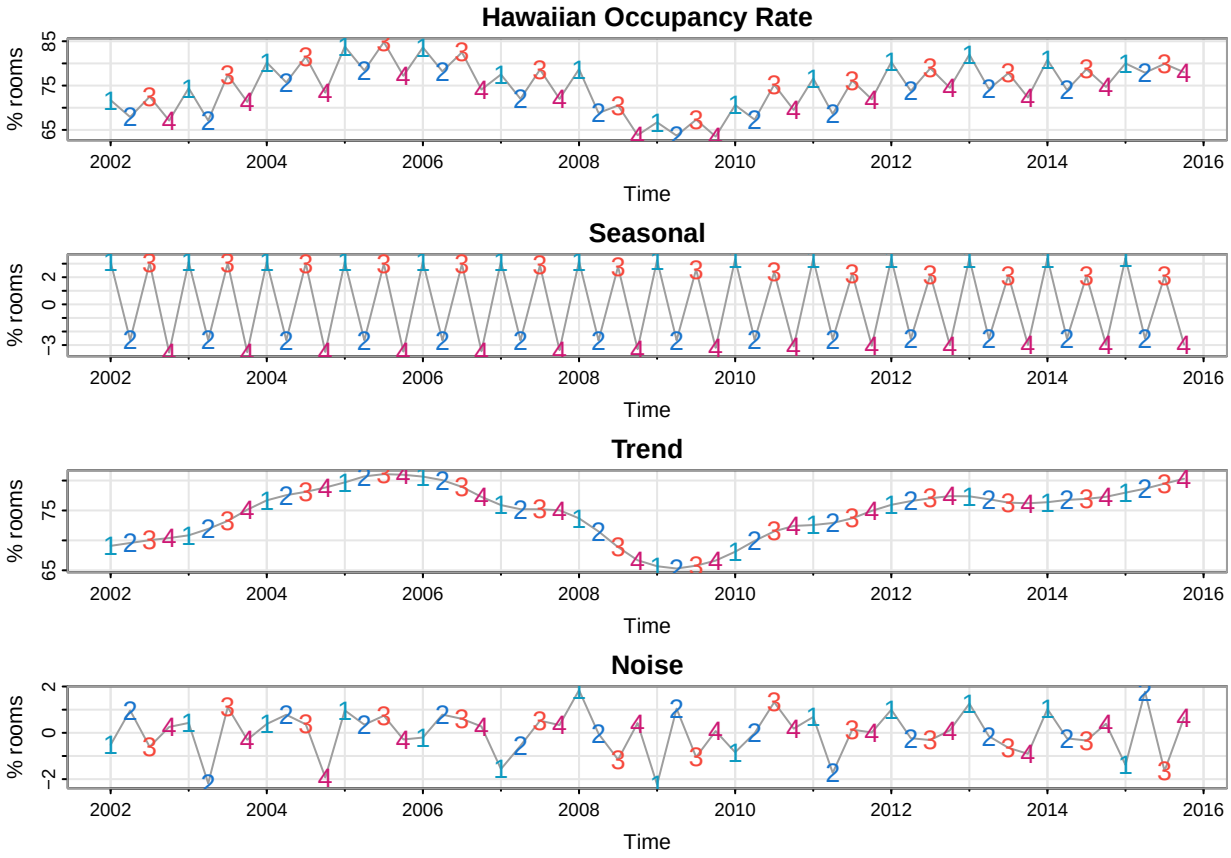

```
plot(stl(x, s.window="per"))
```




```
plot(stl(x, s.window=15))
```




```
culer = c(5, 4, 2, 6)
x <- window(hor, start=2002)
par(mfrow = c(4,1), cex.main=1)
out = stl(x, s.window=15)$time.series
tsplot(x, main='Hawaiian Occupancy Rate', ylab='% rooms', col=8)
text(x, labels=1:4, col=culer, cex=1.25)
tsplot(out[,1], main="Seasonal", ylab='% rooms',col=8)
text(out[,1], labels=1:4, col=culer, cex=1.25)
tsplot(out[,2], main="Trend", ylab='% rooms', col=8)
text(out[,2], labels=1:4, col=culer, cex=1.25)
tsplot(out[,3], main="Noise", ylab='% rooms', col=8)
text(out[,3], labels=1:4, col=culer, cex=1.25)
```


Hawaiian Occupancy Rate
% rooms
65
75
85
2002
2004
2006
2008
2010
2012
2014
2016
Time
Seasonal
% rooms
−3
0
2
Time
Trend
% rooms
65
75
Time
Noise
% rooms
−2
0
2
Time