Time Series Analysis

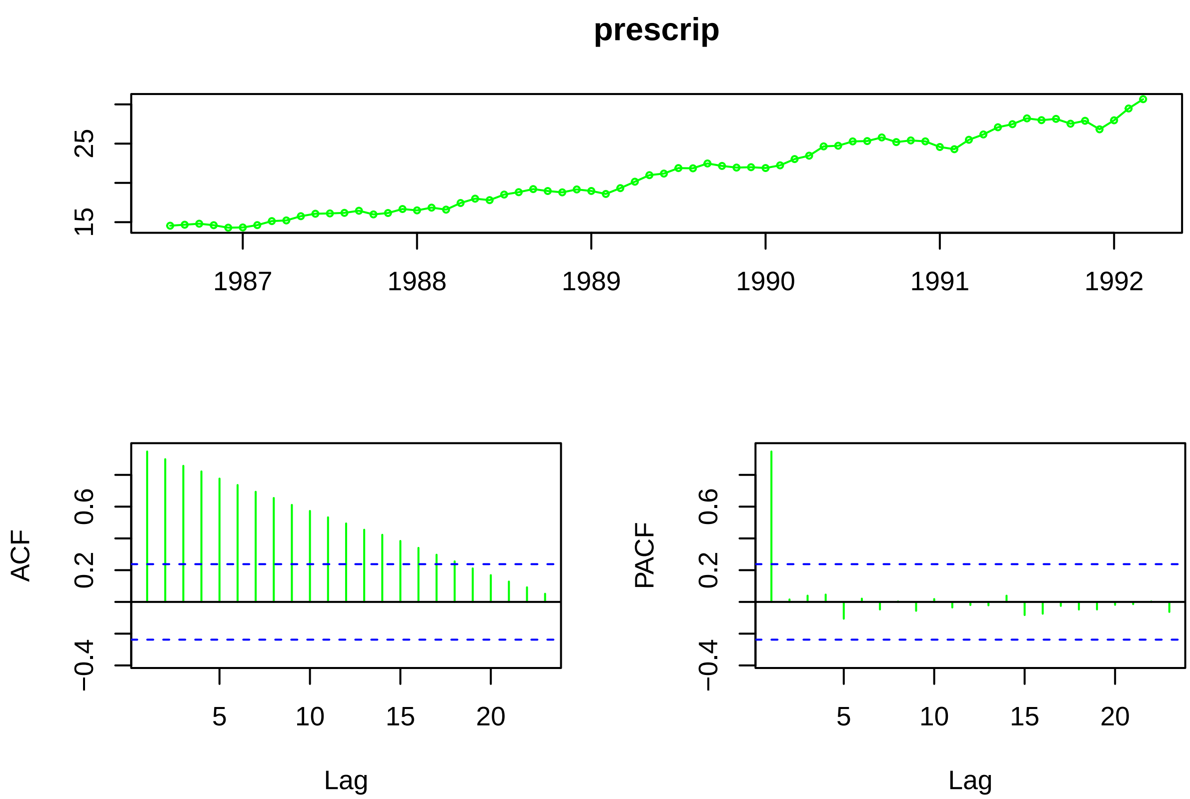
Anisha Mittal

2020-12-4

1. Load Prescriptions data (already available on TSA package)

a. Augmented Diskey - Fuller test

```
data(prescrip)
tsdisplay(prescrip, col = "green")
```




```
adf.test(prescrip, k=0)
```

```
##
##  Augmented Dickey-Fuller Test
```

```
##
## data:  prescrip
## Dickey-Fuller = -2.7968, Lag order = 0, p-value = 0.2515
## alternative hypothesis: stationary
```

Looking at the Augmented Dickey Fuller test, we see that the p-value = 0.2515 > 0.05 (5% significance level) Therefore there does not exist enough evidence to reject the null hypothesis that $\omega = 0$. That is; there is a unit root and the series need to be further differenced to get to stationarity (Alternative Hypothesis) The data plot shows a linear trend. The ACF plot decays to zero at a slow rate which also means that there exists a unit root. The ACF plots does not show any signs of seasonal component.

b. phi(3) test statistics

```
n=length(prescrip)
tt=2:n # convenience vector of time indices
y=diff(prescrip) # first difference of the series
fit=lm(y~tt+prescrip[-n]) # estimate alpha, omega x[t-1], beta
yhat=fitted(fit)
summary(fit)
```

```
##
## Call:
## lm(formula = y ~ tt + prescrip[-n])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.44289 -0.33147  0.01302  0.33358  1.04300
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.12039    1.08902   2.865  0.00563 **
## tt            0.05913    0.01987   2.975  0.00413 **
## prescrip[-n] -0.23663    0.08461  -2.797  0.00681 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4816 on 64 degrees of freedom
## Multiple R-squared:  0.1319, Adjusted R-squared:  0.1048
## F-statistic: 4.862 on 2 and 64 DF,  p-value: 0.01082
```

```
mean(prescrip)
```

```
## [1] 21.0586
```

```
# degrees of freedom of the model = No. of parameters - 1
SSM<-(sum((yhat-mean(y))^2))/2
# degree of freedom of the residuals = No of data points - No. of parameters
SSE<-(sum((y-yhat)^2))/64

Phi3<-(SSM)/(SSE)
Phi3
```

```
## [1] 4.861987
```

c.

```r
A<-ur.df(y,type='trend',lags=0)
summary(A)
```

```
##
## ###############################################
## # Augmented Dickey-Fuller Test Unit Root Test #
## ###############################################
##
## Test regression trend
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + tt)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.45664 -0.37287  0.03616  0.36303  1.03723
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.086679   0.128232   0.676    0.502
## z.lag.1     -0.886861   0.127888  -6.935  2.6e-09 ***
## tt           0.003881   0.003325   1.167    0.248
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5111 on 63 degrees of freedom
## Multiple R-squared:  0.4335, Adjusted R-squared:  0.4155
## F-statistic: 24.11 on 2 and 63 DF,  p-value: 1.68e-08
##
##
## Value of test-statistic is: -6.9347 16.0928 24.1066
##
## Critical values for test statistics:
##       1pct  5pct 10pct
## tau3 -4.04 -3.45 -3.15
## phi2  6.50  4.88  4.16
## phi3  8.73  6.49  5.47
```

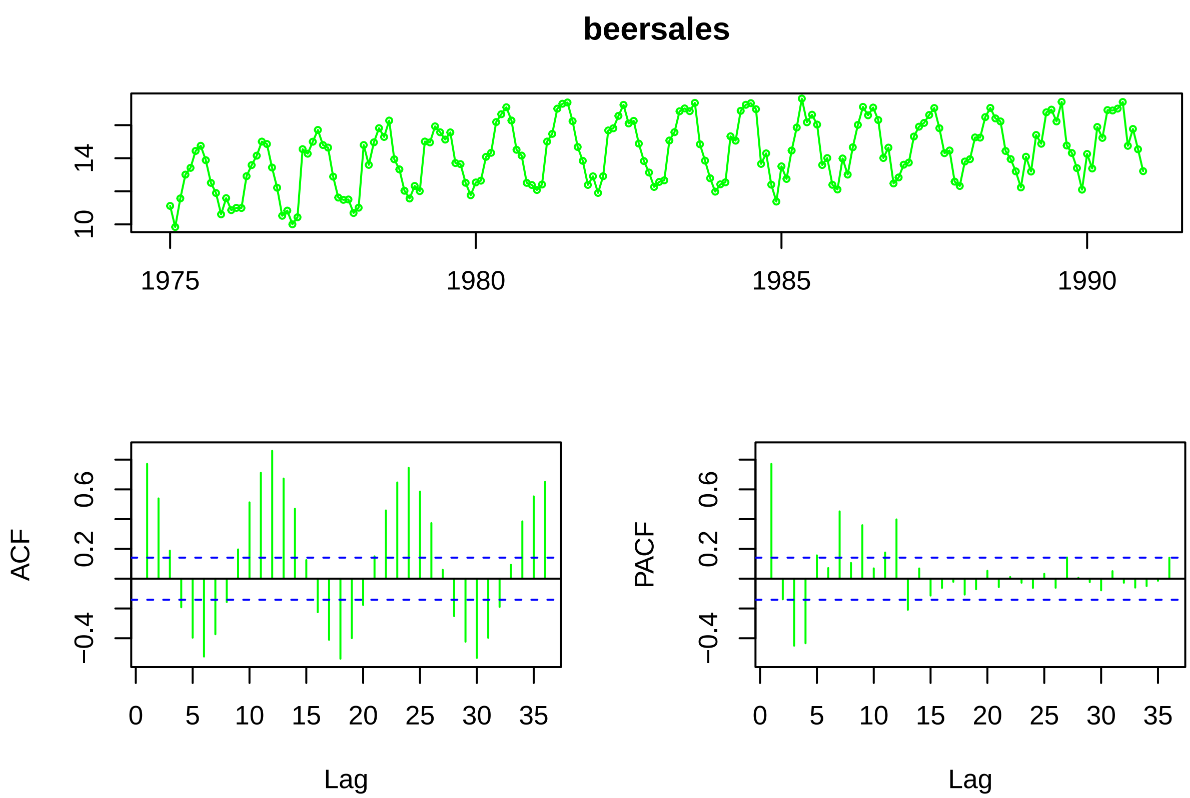
Here we are using the Augmented Dickey-Fuller test for testing the null hypothesis that (alpha, beta, omega) = (aplha, 0, 0) and the alternative hypothesis that (alpha, beta, omega) is not (alpha, 0, 0). We are letting alpha to be estimated freely.

We calculated the value of phi3 as 4.862 in part b and here we observe that critical value for phi 3 is 6.49. So our observed value (of phi 3) 4.862 $<$ the critical value (of phi 3) 6.49 at 5% significance level. Therefore we fail to reject the null hypothesis that (alpha, beta, omega) = (alpha, 0, 0). Which also verifies our claim in part a) that there exist a unit root.

2. Load Beer Sales data (already available on TSA package)

a. Tsdisplay

```
data(beersales)
tsdisplay(beersales, col = "green")
```




```
adf.test(beersales)
```

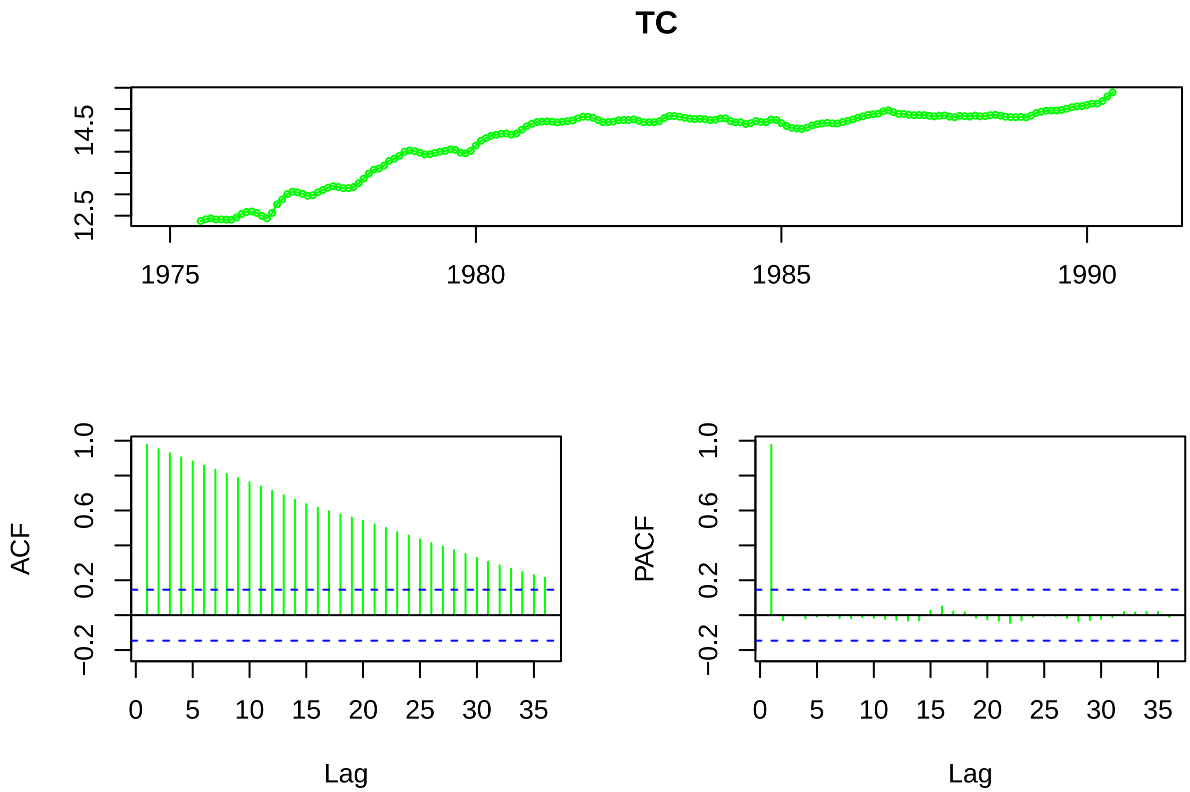
```
## Warning in adf.test(beersales): p-value smaller than printed p-value

##
##  Augmented Dickey-Fuller Test
##
## data:  beersales
## Dickey-Fuller = -9.7734, Lag order = 5, p-value = 0.01
## alternative hypothesis: stationary
```

The time series plot of beersales against time clearly shows that there is a seasonal component due to the regularly seen spikes that occur throughout time. There also seem to be a slight upwards trend as well. We can also see the seasonal component by looking at the ACF plot, it kind of follows the graph of a sine function and we know that the sine function is periodic.

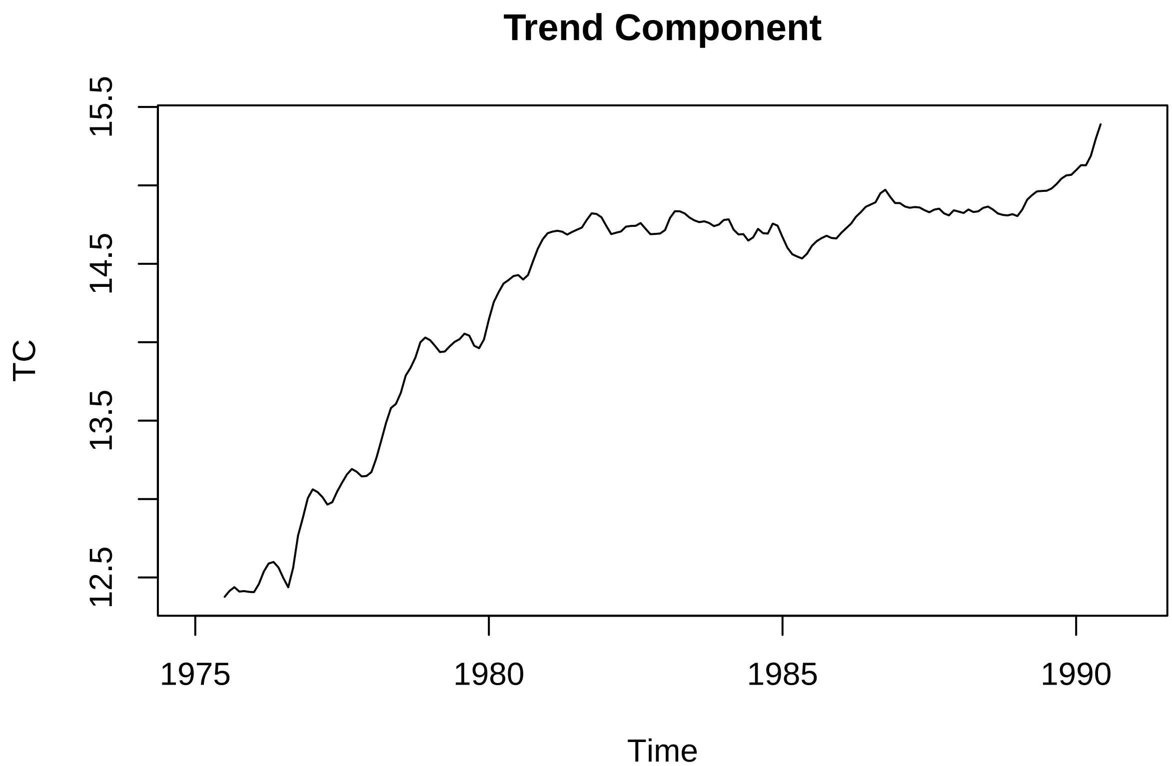
b. Trend estimation and smoothing the series

From the beersales plot in part a, we can see that then spikes occur only once every year. So choosing order 12 for the moving average smoother.

```
TC=ma(beersales,12)
tsdisplay(TC, col = "green")
```

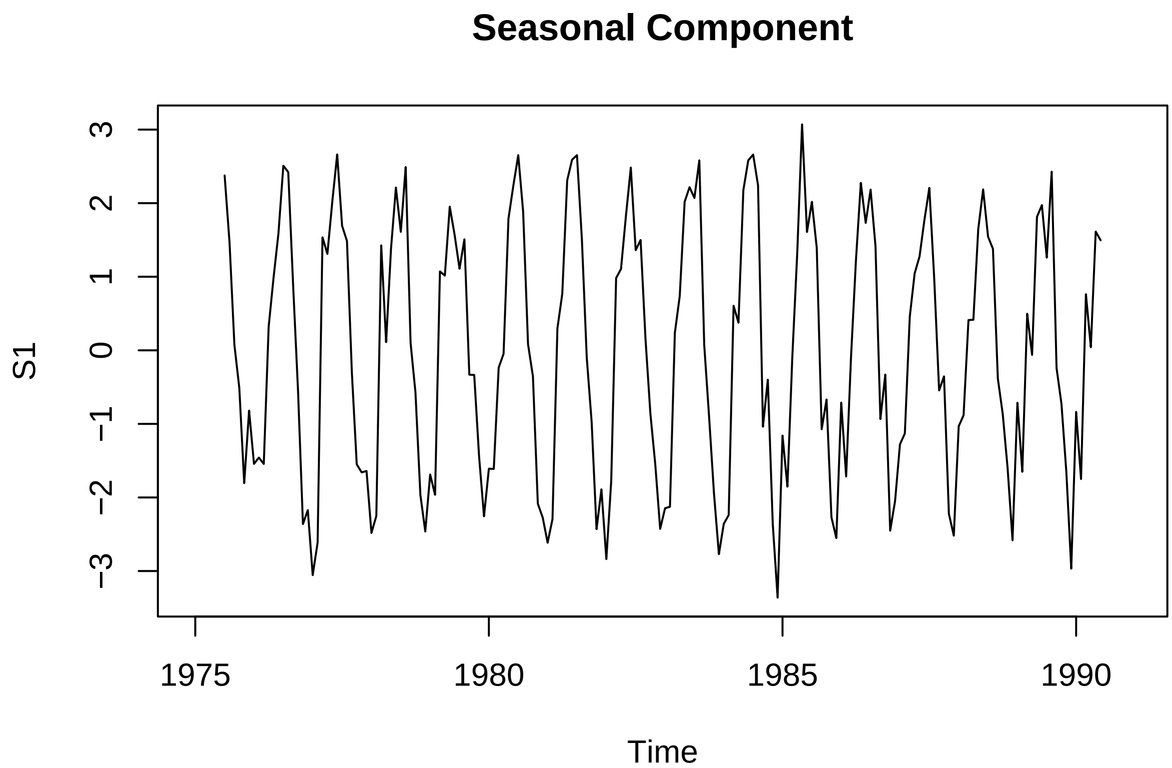



```
plot(TC, main = "Trend Component")
```

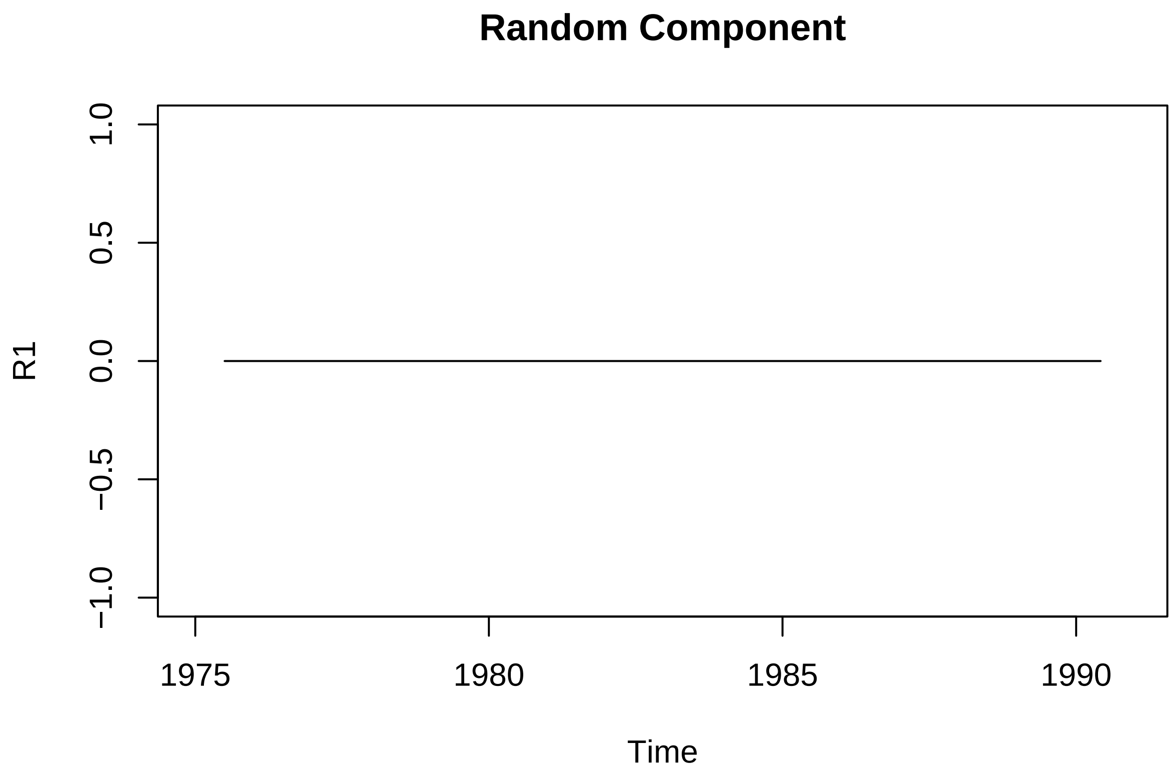


c. Additive Decomposition of the Time Series

```
# find seasonal and random component and plot them against time
S1 = beersales - TC
R1 = beersales - TC - S1
plot(S1, main = "Seasonal Component")
```

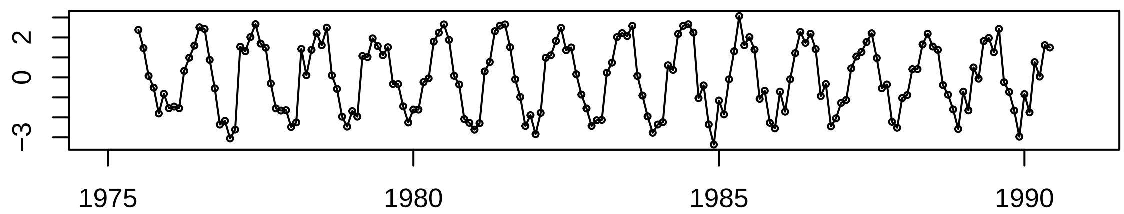


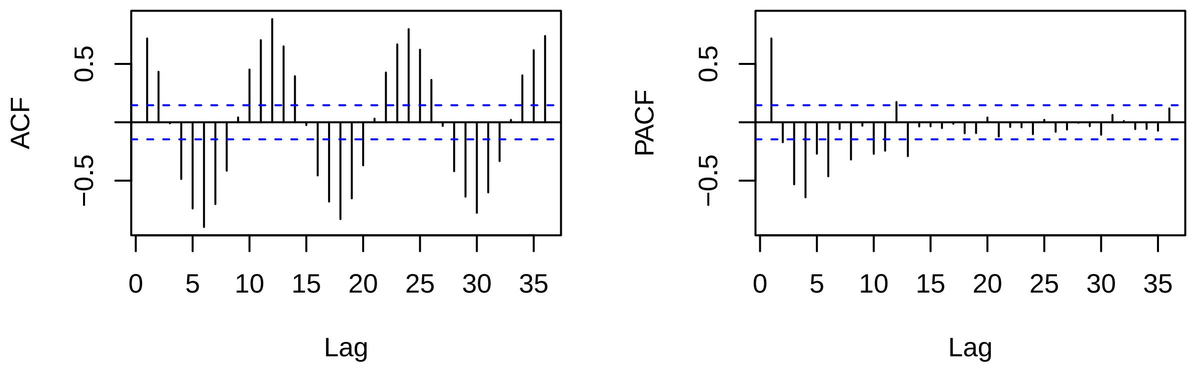

```r
plot(R1, main = "Random Component")
```



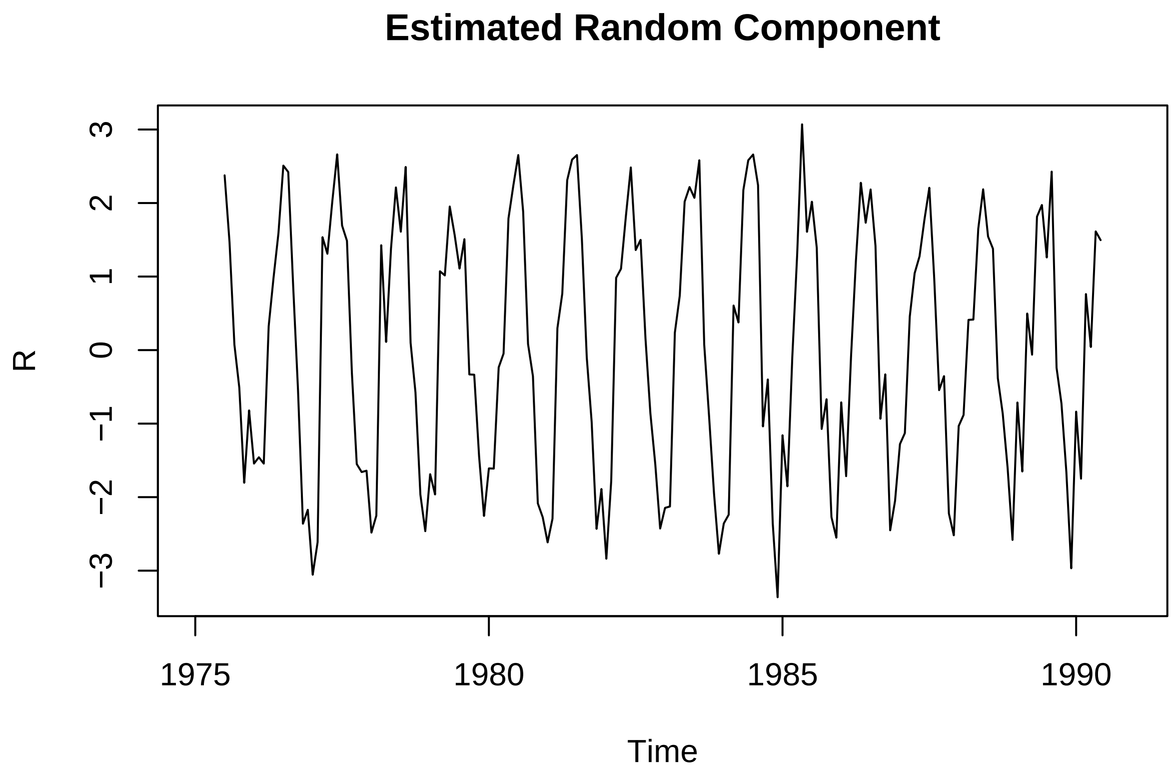

```r
# seasonal component after removing trend component
tsdisplay(beersales-TC ,main = "Series after removing trend component")
```

Series after removing trend component






```
# matrix of seasonal components

pseudo_s=beersales-TC
matrix_s=matrix (pseudo_s, nrow=12)
s=rowMeans (matrix_s, na.rm = TRUE)
srep=rep(length(beersales)/12)
S1 = srep-mean(srep)

# estimate the random component

R=beersales-TC-S1
plot(R, main = "Estimated Random Component")
```




```r
#fitting linear model to trend to forecast
linear_tc=lm (TC~time(beersales))
summary(linear_tc)
```

```
##
## Call:
## lm(formula = TC ~ time(beersales))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.86824 -0.31682 -0.04031  0.34363  0.70449
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -297.29082   13.93976  -21.33   <2e-16 ***
## time(beersales)    0.15714    0.00703   22.35   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4084 on 178 degrees of freedom
##   (12 observations deleted due to missingness)
## Multiple R-squared:  0.7373, Adjusted R-squared:  0.7359
## F-statistic: 499.7 on 1 and 178 DF,  p-value: < 2.2e-16
```

```r
#Verification of the above plots
B=decompose (beersales, type=c("additive"))
plot(B)
```