Chapter 3 Solution

Q1.

For $Y_t = \mu + e_t - e_{t-1}$,

$$
\begin{aligned}
Var(\bar{Y}) &= Var(\frac{1}{n}\sum_{t=1}^{n} Y_t) \\
&= \frac{1}{n^2} Var(\sum_{t=1}^{n} \mu + e_t - e_{t-1}) \\
&= \frac{1}{n^2} Var(e_n - e_0) \\
&= \frac{2}{n^2}\sigma_e^2
\end{aligned}
$$

We know the variance of $\bar{Y}$ for $Y_t = \mu + e_t$ is $\frac{1}{n}\sigma_e^2$, which is much larger. It is because the negative autocorrelation of $Y_t = \mu + e_t - e_{t-1}$ at lag one makes it easier to estimate the process mean when compared with estimating the mean of a white noise process.

Q2.

For $Y_t = \mu + e_t + e_{t-1}$,

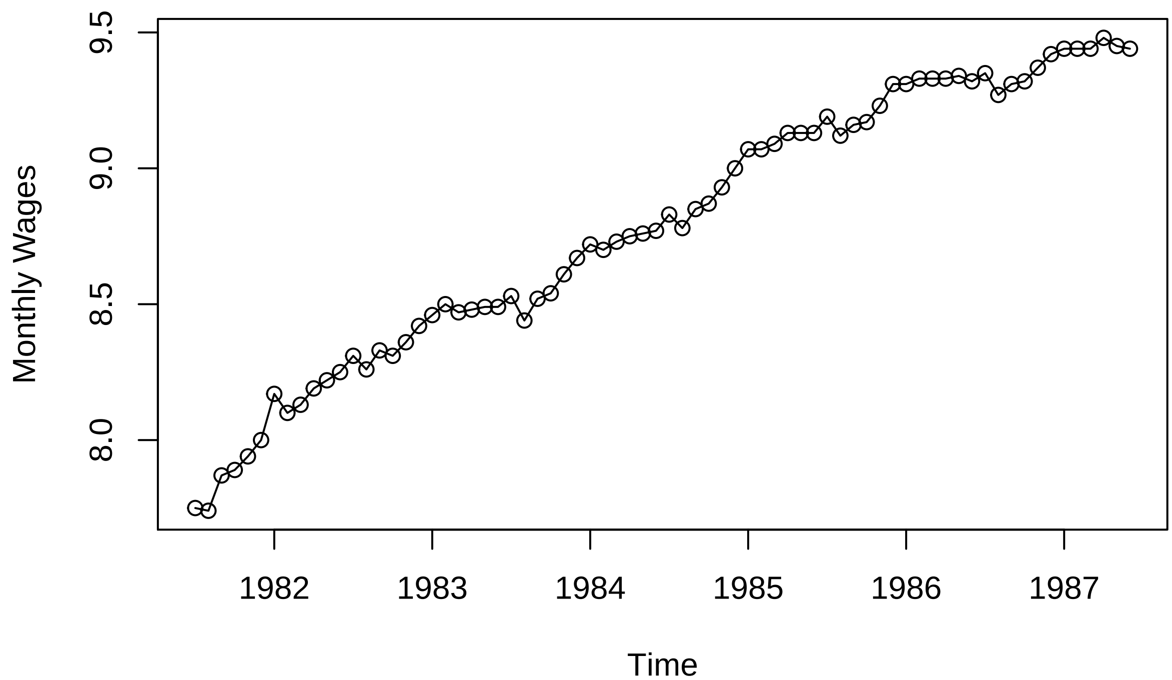
$$
\begin{aligned}
Var(\bar{Y}) &= Var(\frac{1}{n}\sum_{t=1}^{n} Y_t) \\
&= \frac{1}{n^2} Var(\sum_{t=1}^{n} \mu + e_t + e_{t-1}) \\
&= \frac{1}{n^2} Var(e_n + 2(e_{n-1} + ... + e_1) + e_0) \\
&= \frac{4n-2}{n^2}\sigma_e^2
\end{aligned}
$$

The positive autocorrelation at lag one makes it more difficult to estimate the process mean compared with estimating the mean of a white noise process.

Q3.

a.

```
data(wages)
plot(wages,xlab = 'Time', ylab='Monthly Wages',type='o')
```



This plot shows a strong increasing "trend" perhaps linear or curved.

b.

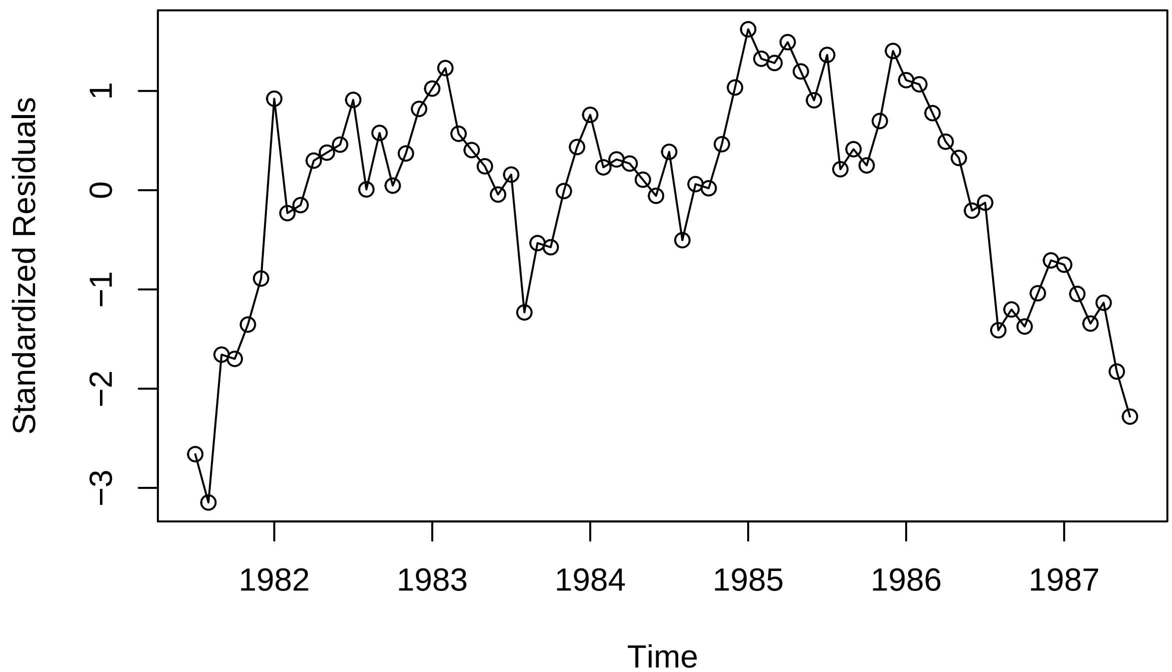
```
wages.lm=lm(wages~time(wages)); summary(wages.lm); y=rstudent(wages.lm)
```

```
##
## Call:
## lm(formula = wages ~ time(wages))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.23828 -0.04981  0.01942  0.05845  0.13136
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5.490e+02  1.115e+01  -49.24   <2e-16 ***
## time(wages)  2.811e-01  5.618e-03   50.03   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.08257 on 70 degrees of freedom
## Multiple R-squared:  0.9728, Adjusted R-squared:  0.9724
## F-statistic:  2503 on 1 and 70 DF,  p-value: < 2.2e-16
```

With a multiple R-squared of 97% and highly significant regression coefficients, it "appears" as if we might have an excellent model.

c.

```r
plot(y,x=as.vector(time(wages)), xlab='Time',ylab='Standardized Residuals',type='o')
```



This plot does not look "random" at all. It has, generally, an upside down U shape and suggests that perhaps we should try a quadratic fit.

d.

```r
wages.lm2=lm(wages~time(wages)+I(time(wages)^2))
summary(wages.lm2); y=rstudent(wages.lm)
```

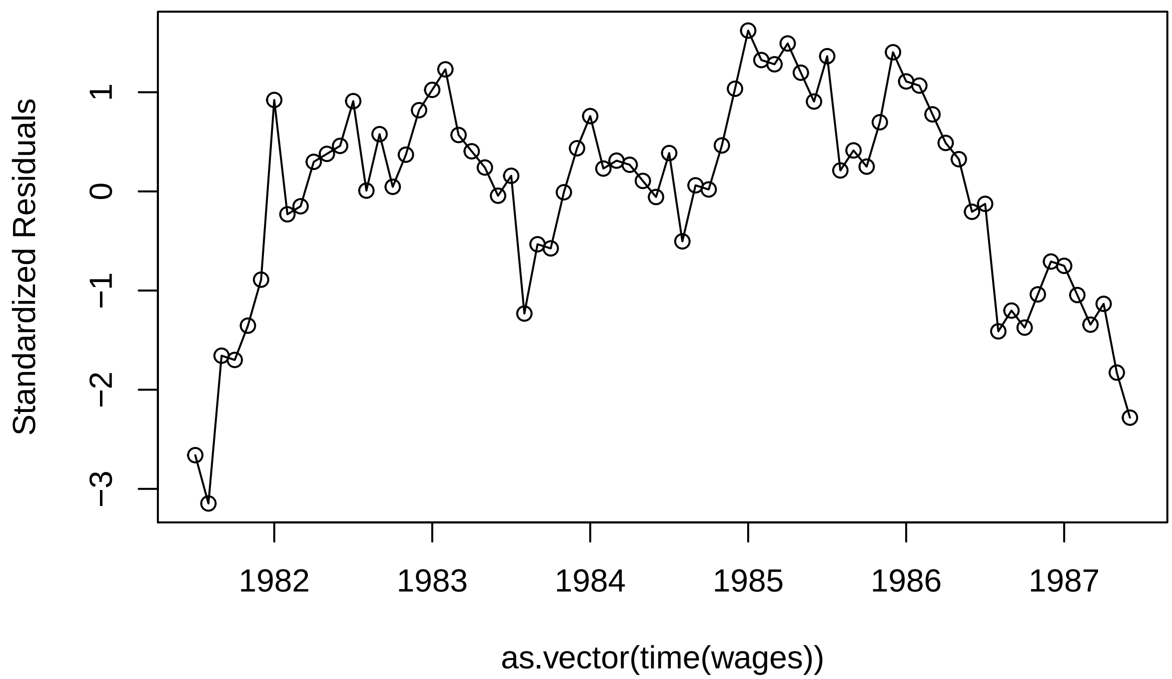
```
##
## Call:
## lm(formula = wages ~ time(wages) + I(time(wages)^2))
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.148318 -0.041440  0.001563  0.050089  0.139839
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -8.495e+04  1.019e+04  -8.336 4.87e-12 ***
## time(wages)       8.534e+01  1.027e+01   8.309 5.44e-12 ***
## I(time(wages)^2) -2.143e-02  2.588e-03  -8.282 6.10e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.05889 on 69 degrees of freedom
## Multiple R-squared:  0.9864, Adjusted R-squared:  0.986
## F-statistic:  2494 on 2 and 69 DF,  p-value: < 2.2e-16
```

Again, based on the regression summary and a 99% R-squared, it "appears" as if we might have an excellent model.

e.

```
plot(y, x=as.vector(time(wages)), ylab = 'Standardized Residuals', type='o')
```

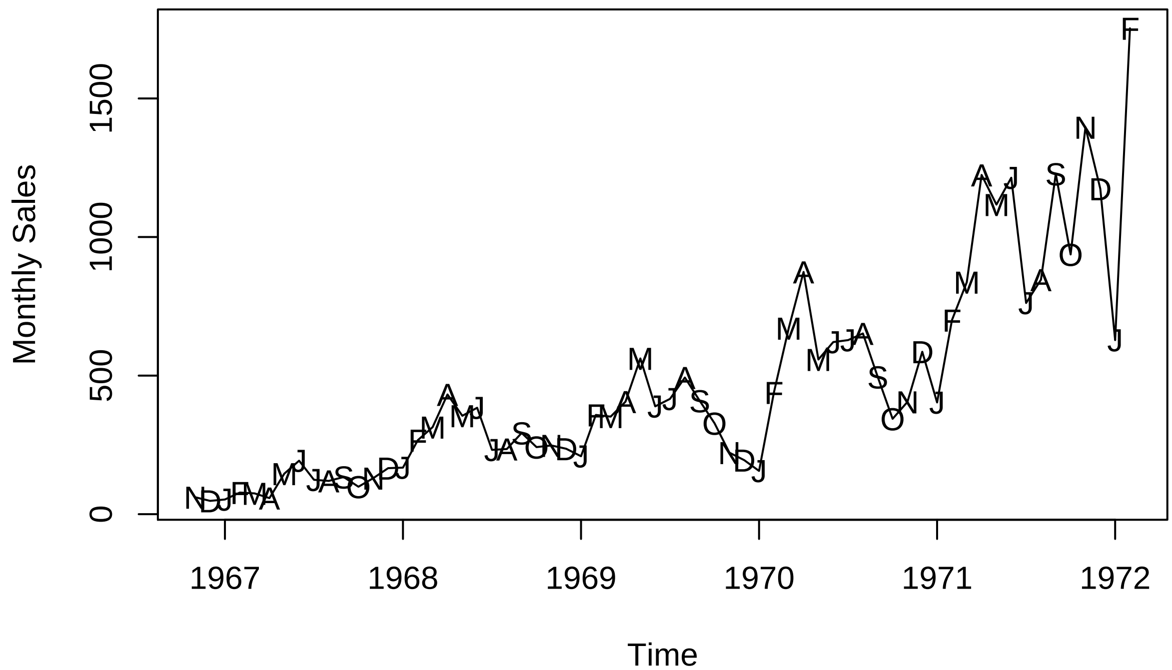


This plot does not look "random" either. It hangs together too much – it is too smooth.

Q4.

a.

```
data(winnebago); plot(winnebago,ylab='Monthly Sales',type='l')
points(y=winnebago,x=time(winnebago), pch=as.vector(season(winnebago)))
```



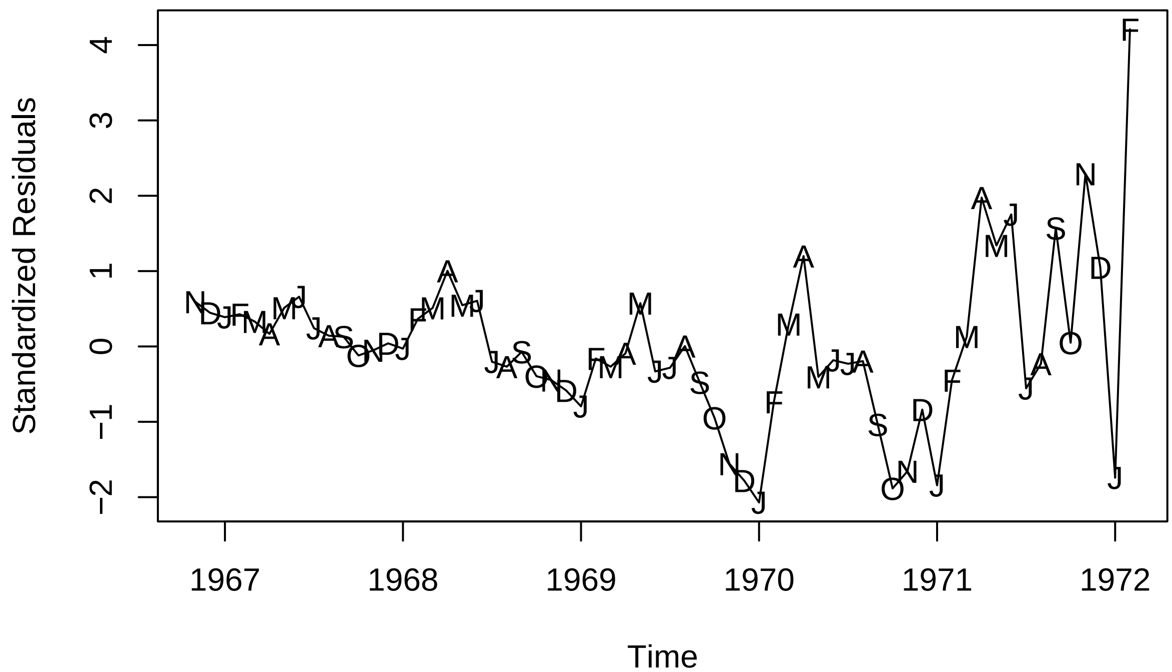
As we would expect with recreational vehicles in the U.S., there is substantial seasonality in the series. However, there is also a general upward "trend" with increasing variation at the higher levels of the series.

b.

```
winnebago.lm=lm(winnebago~time(winnebago)); summary(winnebago.lm)
```

```
##
## Call:
## lm(formula = winnebago ~ time(winnebago))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -419.58  -93.13  -12.78   94.96  759.21
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -394885.68   33539.77  -11.77   <2e-16 ***
## time(winnebago)     200.74      17.03   11.79   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 209.7 on 62 degrees of freedom
## Multiple R-squared:  0.6915, Adjusted R-squared:  0.6865
## F-statistic: 138.9 on 1 and 62 DF,  p-value: < 2.2e-16
```

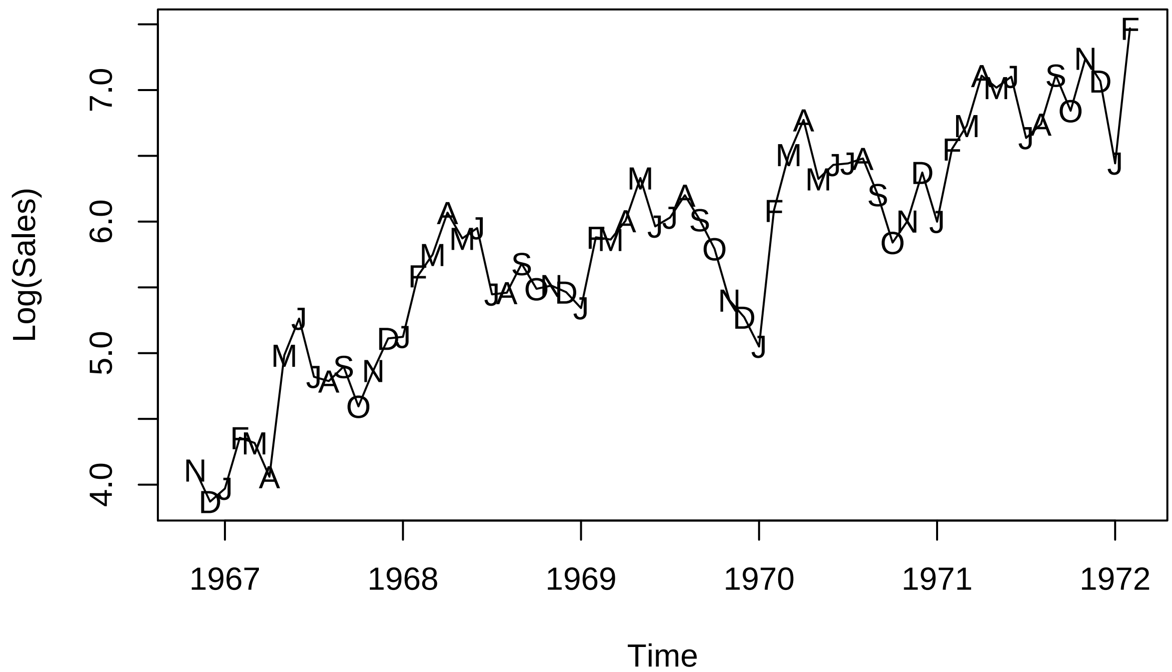
```r
plot(y=rstudent(winnebago.lm),x=as.vector(time(winnebago)),type='l', xlab='Time',ylab='Standardized Resi
points(y=rstudent(winnebago.lm),x=as.vector(time(winnebago)), pch=as.vector(season(winnebago)))
```



Although the “trend” has been removed, this clearly is not an acceptable model and we move on.

c.

```r
plot(log(winnebago), xlab='Time', ylab='Log(Sales)',type='l')
points(y=log(winnebago), x=time(winnebago),pch=as.vector(season(winnebago)))
```



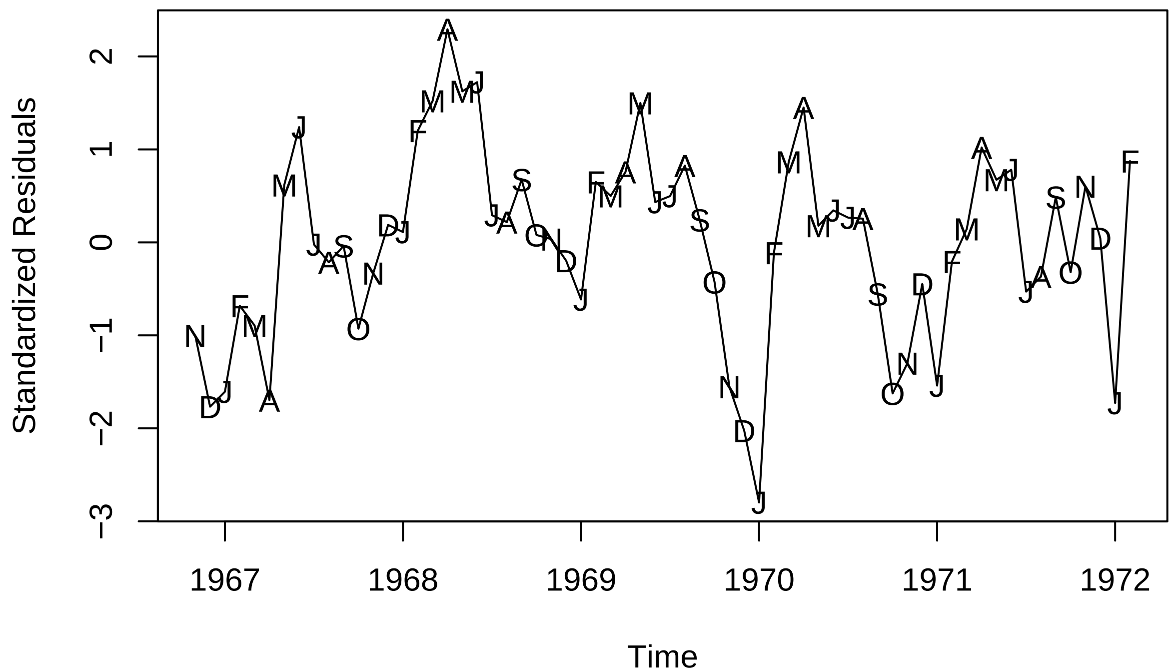
In this we see that the seasonality is still present but that now the upward trend is accompanied by much more equal variation around that trend.

d.

```
logwinnebago.lm=lm(log(winnebago)~time(log(winnebago))); summary(logwinnebago.lm)
```

```
##
## Call:
## lm(formula = log(winnebago) ~ time(log(winnebago)))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.03669 -0.20823  0.04995  0.25662  0.86223
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -984.93878   62.99472  -15.63   <2e-16 ***
## time(log(winnebago))    0.50306    0.03199   15.73   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3939 on 62 degrees of freedom
## Multiple R-squared:  0.7996, Adjusted R-squared:  0.7964
## F-statistic: 247.4 on 1 and 62 DF,  p-value: < 2.2e-16
```

```
plot(y=rstudent(logwinnebago.lm),x=as.vector(time(winnebago)),type='l', xlab='Time', ylab='Standardized
points(y=rstudent(logwinnebago.lm),x=as.vector(time(winnebago)), pch=as.vector(season(winnebago)))
```



The residual plot looks much more acceptable now but we still need to model the seasonality.

e.

```
month.=season(winnebago)
logwinnebago.lm2=lm(log(winnebago)~month.+time(log(winnebago)));summary(logwinnebago.lm2)
```

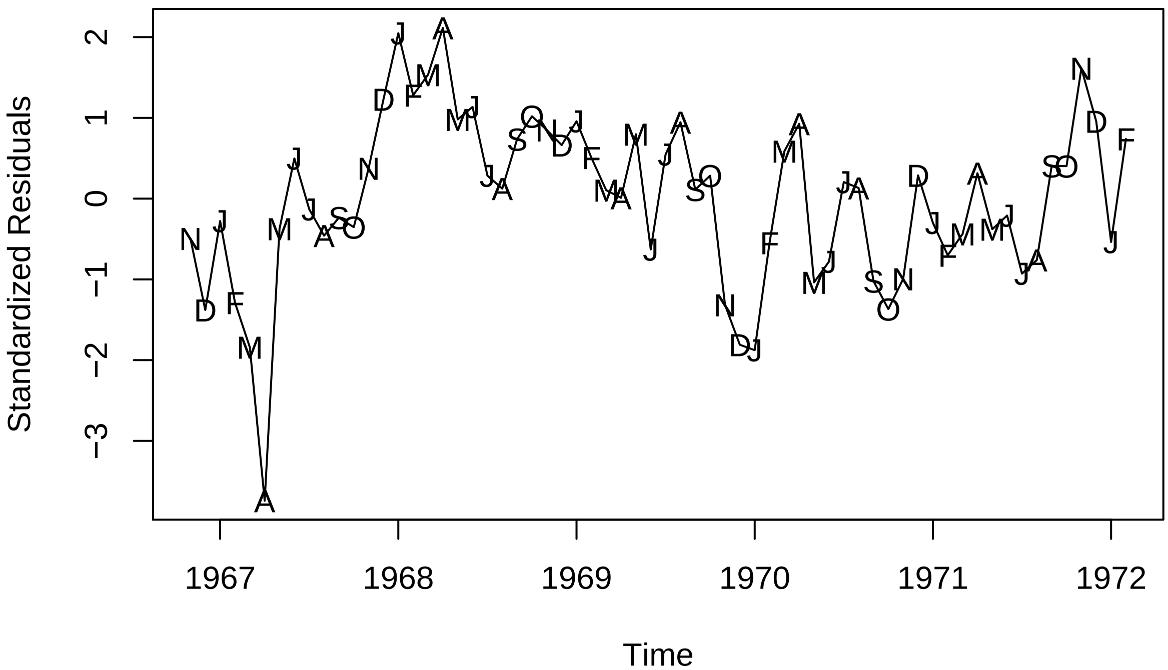
```
##
## Call:
## lm(formula = log(winnebago) ~ month. + time(log(winnebago)))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.92501 -0.16328  0.03344  0.20757  0.57388
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)            -997.33061   50.63995 -19.695  < 2e-16 ***
## month.February            0.62445    0.18182   3.434 0.001188 **
## month.March               0.68220    0.19088   3.574 0.000779 ***
## month.April               0.80959    0.19079   4.243 9.30e-05 ***
## month.May                 0.86953    0.19073   4.559 3.25e-05 ***
## month.June                0.86309    0.19070   4.526 3.63e-05 ***
```

```
## month.July              0.55392    0.19069   2.905 0.005420 **
## month.August            0.56989    0.19070   2.988 0.004305 **
## month.September         0.57572    0.19073   3.018 0.003960 **
## month.October           0.26349    0.19079   1.381 0.173300
## month.November          0.28682    0.18186   1.577 0.120946
## month.December          0.24802    0.18182   1.364 0.178532
## time(log(winnebago))    0.50909    0.02571  19.800  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3149 on 51 degrees of freedom
## Multiple R-squared:  0.8946, Adjusted R-squared:  0.8699
## F-statistic: 36.09 on 12 and 51 DF,  p-value: < 2.2e-16
```

This model explains a large percentage of the variation in sales but, as always, we should also look at the residuals.

f.

```r
plot(y=rstudent(logwinnebago.lm2),x=as.vector(time(winnebago)),type='l', xlab='Time', ylab='Standardize
points(y=rstudent(logwinnebago.lm2),x=as.vector(time(winnebago)), pch=as.vector(season(winnebago)))
```



This residual plot is the best we have seen for models of this series but perhaps there are better models to be explored later.