Q. 1)

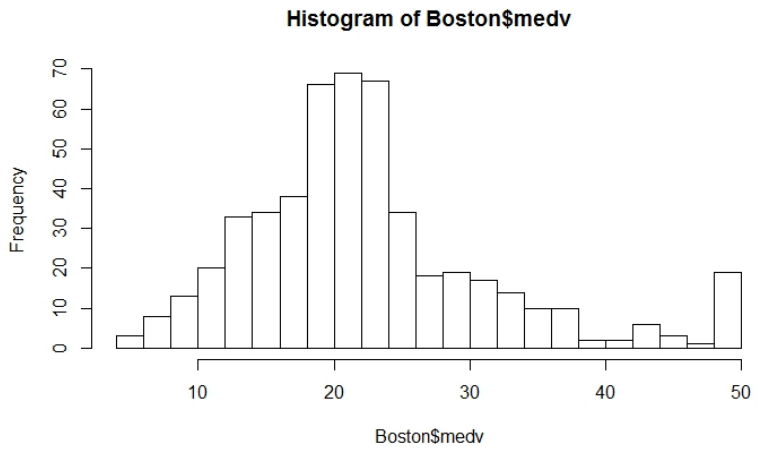
```
> library(MASS)
> data(Boston)
> head(Boston)
     crim zn indus chas   nox    rm  age    dis rad tax ptratio  black lstat medv
1 0.00632 18  2.31    0 0.538 6.575 65.2 4.0900   1 296    15.3 396.90  4.98 24.0
2 0.02731  0  7.07    0 0.469 6.421 78.9 4.9671   2 242    17.8 396.90  9.14 21.6
3 0.02729  0  7.07    0 0.469 7.185 61.1 4.9671   2 242    17.8 392.83  4.03 34.7
4 0.03237  0  2.18    0 0.458 6.998 45.8 6.0622   3 222    18.7 394.63  2.94 33.4
5 0.06905  0  2.18    0 0.458 7.147 54.2 6.0622   3 222    18.7 396.90  5.33 36.2
6 0.02985  0  2.18    0 0.458 6.430 58.7 6.0622   3 222    18.7 394.12  5.21 28.7

> LM1 = lm(medv ~ crim + zn + indus + chas + nox + rm + age + dis + rad +
tax + ptratio + black + lstat, data = Boston)

> summary(LM1)

Call:
lm(formula = medv ~ crim + zn + indus + chas + nox + rm + age +
    dis + rad + tax + ptratio + black + lstat, data = Boston)

Residuals:
    Min      1Q  Median      3Q     Max
-15.595  -2.730  -0.518   1.777  26.199

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.646e+01  5.103e+00   7.144 3.28e-12 ***
crim        -1.080e-01  3.286e-02  -3.287 0.001087 **
zn           4.642e-02  1.373e-02   3.382 0.000778 ***
indus        2.056e-02  6.150e-02   0.334 0.738288
chas         2.687e+00  8.616e-01   3.118 0.001925 **
nox         -1.777e+01  3.820e+00  -4.651 4.25e-06 ***
rm           3.810e+00  4.179e-01   9.116  < 2e-16 ***
age          6.922e-04  1.321e-02   0.052 0.958229
dis         -1.476e+00  1.995e-01  -7.398 6.01e-13 ***
rad          3.060e-01  6.635e-02   4.613 5.07e-06 ***
tax         -1.233e-02  3.760e-03  -3.280 0.001112 **
ptratio     -9.527e-01  1.308e-01  -7.283 1.31e-12 ***
black        9.312e-03  2.686e-03   3.467 0.000573 ***
lstat       -5.248e-01  5.072e-02 -10.347  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.745 on 492 degrees of freedom
Multiple R-squared:  0.7406,   Adjusted R-squared:  0.7338
F-statistic: 108.1 on 13 and 492 DF,  p-value: < 2.2e-16


> AIC(LM1)
[1] 3027.609
> BIC(LM1)
[1] 3091.007

> n1 <- nrow(Boston)

> LM0 = lm(medv ~ 1, data=Boston)
> AIC(LM0)
[1] 3684.48
> AIC(LM0, k = log(n1))
[1] 3692.933
> BIC(LM0)
[1] 3692.933


> fit1 = step(LM1 , direction = "backward")
Start:  AIC=1589.64
```

```
medv ~ crim + zn + indus + chas + nox + rm + age + dis + rad +
    tax + ptratio + black + lstat

          Df Sum of Sq   RSS    AIC
- age      1      0.06 11079 1587.7
- indus    1      2.52 11081 1587.8
<none>                 11079 1589.6
- chas     1    218.97 11298 1597.5
- tax      1    242.26 11321 1598.6
- crim     1    243.22 11322 1598.6
- zn       1    257.49 11336 1599.3
- black    1    270.63 11349 1599.8
- rad      1    479.15 11558 1609.1
- nox      1    487.16 11566 1609.4
- ptratio  1   1194.23 12273 1639.4
- dis      1   1232.41 12311 1641.0
- rm       1   1871.32 12950 1666.6
- lstat    1   2410.84 13490 1687.3

Step:  AIC=1587.65
medv ~ crim + zn + indus + chas + nox + rm + dis + rad + tax +
    ptratio + black + lstat

          Df Sum of Sq   RSS    AIC
- indus    1      2.52 11081 1585.8
<none>                 11079 1587.7
- chas     1    219.91 11299 1595.6
- tax      1    242.24 11321 1596.6
- crim     1    243.20 11322 1596.6
- zn       1    260.32 11339 1597.4
- black    1    272.26 11351 1597.9
- rad      1    481.09 11560 1607.2
- nox      1    520.87 11600 1608.9
- ptratio  1   1200.23 12279 1637.7
- dis      1   1352.26 12431 1643.9
- rm       1   1959.55 13038 1668.0
- lstat    1   2718.88 13798 1696.7

Step:  AIC=1585.76
medv ~ crim + zn + chas + nox + rm + dis + rad + tax + ptratio +
    black + lstat

          Df Sum of Sq   RSS    AIC
<none>                 11081 1585.8
- chas     1    227.21 11309 1594.0
- crim     1    245.37 11327 1594.8
- zn       1    257.82 11339 1595.4
- black    1    270.82 11352 1596.0
- tax      1    273.62 11355 1596.1
- rad      1    500.92 11582 1606.1
- nox      1    541.91 11623 1607.9
- ptratio  1   1206.45 12288 1636.0
- dis      1   1448.94 12530 1645.9
- rm       1   1963.66 13045 1666.3
- lstat    1   2723.48 13805 1695.0

> fit2 = step(LM1, direction = "backward", k=log(n1))
Start:  AIC=1648.81
medv ~ crim + zn + indus + chas + nox + rm + age + dis + rad +
    tax + ptratio + black + lstat

          Df Sum of Sq   RSS    AIC
- age      1      0.06 11079 1642.6
- indus    1      2.52 11081 1642.7
<none>                 11079 1648.8
- chas     1    218.97 11298 1652.5
- tax      1    242.26 11321 1653.5
- crim     1    243.22 11322 1653.6
- zn       1    257.49 11336 1654.2
```

```
- black    1    270.63 11349 1654.8
- rad      1    479.15 11558 1664.0
- nox      1    487.16 11566 1664.4
- ptratio  1   1194.23 12273 1694.4
- dis      1   1232.41 12311 1696.0
- rm       1   1871.32 12950 1721.6
- lstat    1   2410.84 13490 1742.2

Step:  AIC=1642.59
medv ~ crim + zn + indus + chas + nox + rm + dis + rad + tax +
    ptratio + black + lstat

          Df Sum of Sq   RSS    AIC
- indus    1      2.52 11081 1636.5
<none>                 11079 1642.6
- chas     1    219.91 11299 1646.3
- tax      1    242.24 11321 1647.3
- crim     1    243.20 11322 1647.3
- zn       1    260.32 11339 1648.1
- black    1    272.26 11351 1648.7
- rad      1    481.09 11560 1657.9
- nox      1    520.87 11600 1659.6
- ptratio  1   1200.23 12279 1688.4
- dis      1   1352.26 12431 1694.6
- rm       1   1959.55 13038 1718.8
- lstat    1   2718.88 13798 1747.4

Step:  AIC=1636.48
medv ~ crim + zn + chas + nox + rm + dis + rad + tax + ptratio +
    black + lstat

          Df Sum of Sq   RSS    AIC
<none>                 11081 1636.5
- chas     1    227.21 11309 1640.5
- crim     1    245.37 11327 1641.3
- zn       1    257.82 11339 1641.9
- black    1    270.82 11352 1642.5
- tax      1    273.62 11355 1642.6
- rad      1    500.92 11582 1652.6
- nox      1    541.91 11623 1654.4
- ptratio  1   1206.45 12288 1682.5
- dis      1   1448.94 12530 1692.4
- rm       1   1963.66 13045 1712.8
- lstat    1   2723.48 13805 1741.5
```

AIC and BIC both selected the same eleven variables and eliminated 2 variables.

```
> hist(Boston$medv,breaks=20)
```



It is skewed, so we can use "log(medv)" instead of "medv" as follows:

```
> LM2 = lm(log(medv) ~ crim + zn + indus + chas + nox + rm + age + dis + rad +
+              tax + ptratio + black + lstat, data = Boston)

> summary(LM2)

Call:
lm(formula = log(medv) ~ crim + zn + indus + chas + nox + rm +
    age + dis + rad + tax + ptratio + black + lstat, data = Boston)

Residuals:
     Min       1Q   Median       3Q      Max
-0.73361 -0.09747 -0.01657  0.09629  0.86435

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.1020423  0.2042726  20.081  < 2e-16 ***
crim        -0.0102715  0.0013155  -7.808 3.52e-14 ***
zn           0.0011725  0.0005495   2.134 0.033349 *
indus        0.0024668  0.0024614   1.002 0.316755
chas         0.1008876  0.0344859   2.925 0.003598 **
nox         -0.7783993  0.1528902  -5.091 5.07e-07 ***
rm           0.0908331  0.0167280   5.430 8.87e-08 ***
age          0.0002106  0.0005287   0.398 0.690567
dis         -0.0490873  0.0079834  -6.149 1.62e-09 ***
rad          0.0142673  0.0026556   5.373 1.20e-07 ***
tax         -0.0006258  0.0001505  -4.157 3.80e-05 ***
ptratio     -0.0382715  0.0052365  -7.309 1.10e-12 ***
black        0.0004136  0.0001075   3.847 0.000135 ***
lstat       -0.0290355  0.0020299 -14.304  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1899 on 492 degrees of freedom
Multiple R-squared:  0.7896,   Adjusted R-squared:  0.7841
F-statistic: 142.1 on 13 and 492 DF,  p-value: < 2.2e-16

> fit12 = step(LM2 , direction = "backward")
Start:  AIC=-1667.19
log(medv) ~ crim + zn + indus + chas + nox + rm + age + dis +
    rad + tax + ptratio + black + lstat

          Df Sum of Sq    RSS     AIC
- age      1    0.0057 17.755 -1669.0
- indus    1    0.0362 17.786 -1668.2
<none>                17.749 -1667.2
- zn       1    0.1643 17.914 -1664.5
- chas     1    0.3088 18.058 -1660.5
- black    1    0.5339 18.283 -1654.2
- tax      1    0.6235 18.373 -1651.7
- nox      1    0.9351 18.684 -1643.2
- rad      1    1.0413 18.791 -1640.3
- rm       1    1.0637 18.813 -1639.7
- dis      1    1.3639 19.113 -1631.7
- ptratio  1    1.9270 19.676 -1617.0
- crim     1    2.1995 19.949 -1610.1
- lstat    1    7.3809 25.130 -1493.2

Step:  AIC=-1669.03
```

```
log(medv) ~ crim + zn + indus + chas + nox + rm + dis + rad +
    tax + ptratio + black + lstat

          Df Sum of Sq    RSS     AIC
- indus    1    0.0363 17.791 -1670.0
<none>                 17.755 -1669.0
- zn       1    0.1593 17.914 -1666.5
- chas     1    0.3138 18.069 -1662.2
- black    1    0.5431 18.298 -1655.8
- tax      1    0.6205 18.376 -1653.7
- nox      1    0.9645 18.720 -1644.3
- rad      1    1.0356 18.791 -1642.3
- rm       1    1.1452 18.900 -1639.4
- dis      1    1.5471 19.302 -1628.8
- ptratio  1    1.9224 19.677 -1619.0
- crim     1    2.1988 19.954 -1612.0
- lstat    1    8.1949 25.950 -1479.0

Step:  AIC=-1670
log(medv) ~ crim + zn + chas + nox + rm + dis + rad + tax + ptratio +
    black + lstat

          Df Sum of Sq    RSS     AIC
<none>                 17.791 -1670.0
- zn       1    0.1451 17.936 -1667.9
- chas     1    0.3399 18.131 -1662.4
- black    1    0.5344 18.326 -1657.0
- tax      1    0.6139 18.405 -1654.8
- nox      1    0.9350 18.726 -1646.1
- rad      1    1.0088 18.800 -1644.1
- rm       1    1.1171 18.909 -1641.2
- dis      1    1.7385 19.530 -1624.8
- ptratio  1    1.8862 19.678 -1621.0
- crim     1    2.2229 20.014 -1612.4
- lstat    1    8.1604 25.952 -1481.0

> fit22 = step(LM2, direction = "backward", k=log(n1))
Start:  AIC=-1608.02
log(medv) ~ crim + zn + indus + chas + nox + rm + age + dis +
    rad + tax + ptratio + black + lstat

          Df Sum of Sq    RSS     AIC
- age      1    0.0057 17.755 -1614.1
- indus    1    0.0362 17.786 -1613.2
- zn       1    0.1643 17.914 -1609.6
<none>                 17.749 -1608.0
- chas     1    0.3088 18.058 -1605.5
- black    1    0.5339 18.283 -1599.2
- tax      1    0.6235 18.373 -1596.8
- nox      1    0.9351 18.684 -1588.3
- rad      1    1.0413 18.791 -1585.4
- rm       1    1.0637 18.813 -1584.8
- dis      1    1.3639 19.113 -1576.8
- ptratio  1    1.9270 19.676 -1562.1
- crim     1    2.1995 19.949 -1555.1
- lstat    1    7.3809 25.130 -1438.3

Step:  AIC=-1614.09
log(medv) ~ crim + zn + indus + chas + nox + rm + dis + rad +
    tax + ptratio + black + lstat
```

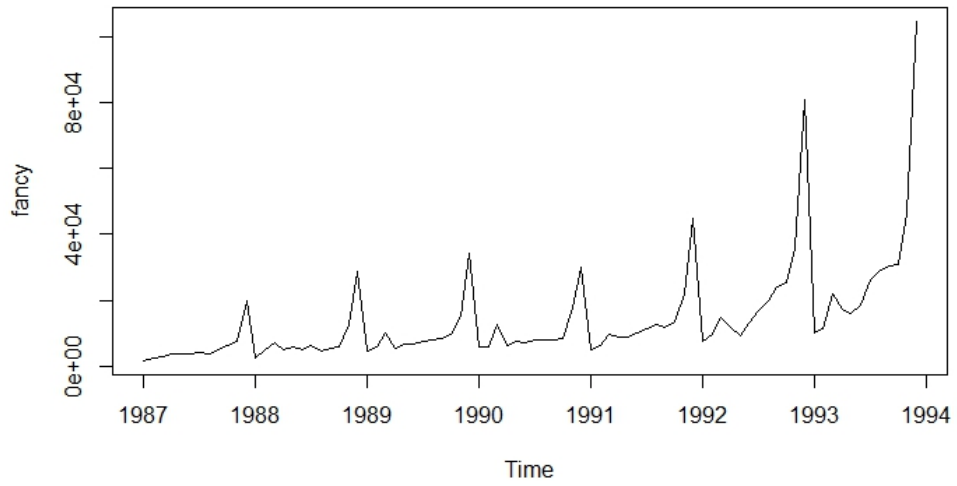
```
          Df Sum of Sq    RSS     AIC
- indus    1    0.0363 17.791 -1619.3
- zn       1    0.1593 17.914 -1615.8
<none>                  17.755 -1614.1
- chas     1    0.3138 18.069 -1611.5
- black    1    0.5431 18.298 -1605.1
- tax      1    0.6205 18.376 -1602.9
- nox      1    0.9645 18.720 -1593.5
- rad      1    1.0356 18.791 -1591.6
- rm       1    1.1452 18.900 -1588.7
- dis      1    1.5471 19.302 -1578.0
- ptratio  1    1.9224 19.677 -1568.3
- crim     1    2.1988 19.954 -1561.2
- lstat    1    8.1949 25.950 -1428.3

Step:  AIC=-1619.28
log(medv) ~ crim + zn + chas + nox + rm + dis + rad + tax + ptratio +
    black + lstat

          Df Sum of Sq    RSS     AIC
- zn       1    0.1451 17.936 -1621.4
<none>                  17.791 -1619.3
- chas     1    0.3399 18.131 -1615.9
- black    1    0.5344 18.326 -1610.5
- tax      1    0.6139 18.405 -1608.3
- nox      1    0.9350 18.726 -1599.6
- rad      1    1.0088 18.800 -1597.6
- rm       1    1.1171 18.909 -1594.7
- dis      1    1.7385 19.530 -1578.3
- ptratio  1    1.8862 19.678 -1574.5
- crim     1    2.2229 20.014 -1565.9
- lstat    1    8.1604 25.952 -1434.5

Step:  AIC=-1621.4
log(medv) ~ crim + chas + nox + rm + dis + rad + tax + ptratio +
    black + lstat

          Df Sum of Sq    RSS     AIC
<none>                  17.936 -1621.4
- chas     1    0.3388 18.275 -1618.2
- tax      1    0.5229 18.459 -1613.1
- black    1    0.5386 18.475 -1612.7
- rad      1    0.9601 18.897 -1601.2
- nox      1    1.0250 18.961 -1599.5
- rm       1    1.2650 19.201 -1593.1
- dis      1    1.6967 19.633 -1581.9
- crim     1    2.1377 20.074 -1570.7
- ptratio  1    2.5632 20.500 -1560.0
- lstat    1    8.1516 26.088 -1438.1
```

Here, the results of AIC and BIC are different. Using BIC we eliminated one more variable “zn”.

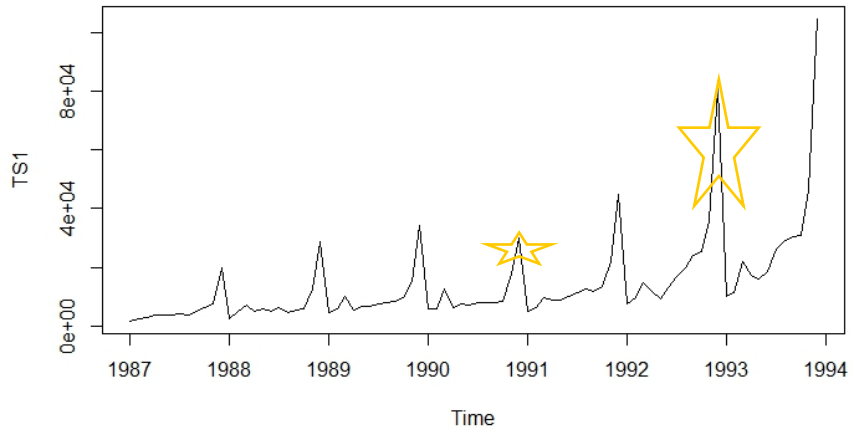
Q. 2)

```
> library(fpp)
> data(fancy)
> head(fancy)
[1] 1664.81 2397.53 2840.71 3547.29 3752.96 3714.74

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

```
> plot(fancy)
```

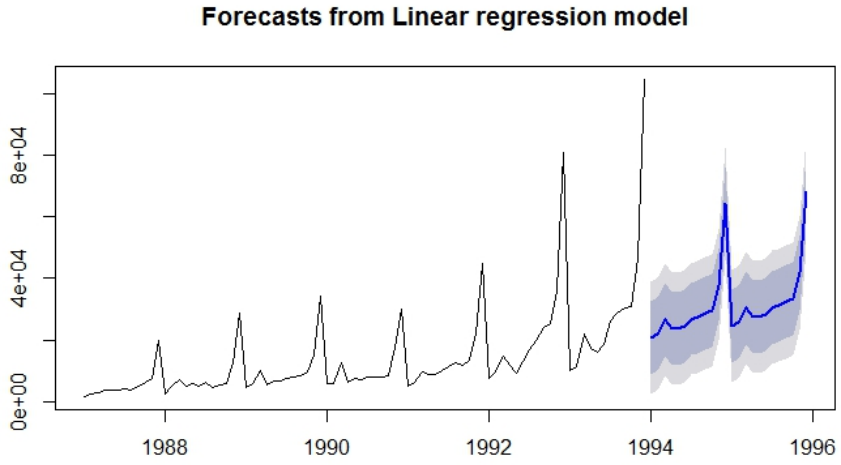


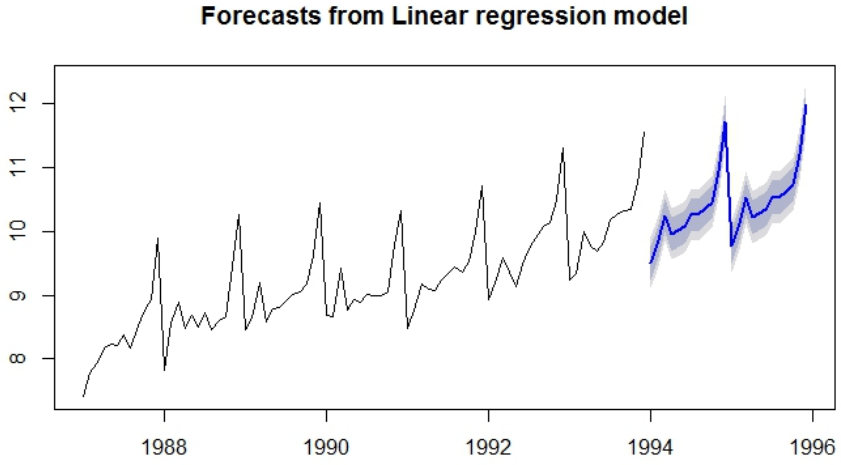

```
> TS1 = ts(fancy, start=c(1987,1), frequency=12)
> plot(TS1)
```




```
> fit11 <- tslm(TS1 ~ trend + season)
> sale11 = ts(fancy, start=c(1987,1), frequency=12)

> fit12 = tslm(log(sale11) ~ trend + season)
> summary(fit12)

Call:
tslm(formula = log(sale11) ~ trend + season)

Residuals:
     Min       1Q   Median       3Q      Max
-0.41644 -0.12619  0.00608  0.11389  0.38567

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 7.6058604  0.0768740  98.939  < 2e-16 ***
trend       0.0223930  0.0008448  26.508  < 2e-16 ***
season2     0.2510437  0.0993278   2.527 0.013718 *
season3     0.6952066  0.0993386   6.998 1.18e-09 ***
season4     0.3829341  0.0993565   3.854 0.000252 ***
```

```
season5     0.4079944  0.0993817   4.105 0.000106 ***
season6     0.4469625  0.0994140   4.496 2.63e-05 ***
season7     0.6082156  0.0994534   6.116 4.69e-08 ***
season8     0.5853524  0.0995001   5.883 1.21e-07 ***
season9     0.6663446  0.0995538   6.693 4.27e-09 ***
season10    0.7440336  0.0996148   7.469 1.61e-10 ***
season11    1.2030164  0.0996828  12.068  < 2e-16 ***
season12    1.9581366  0.0997579  19.629  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1858 on 71 degrees of freedom
Multiple R-squared:  0.9527,   Adjusted R-squared:  0.9447
F-statistic: 119.1 on 12 and 71 DF,  p-value: < 2.2e-16

> par(mfrow=c(1,2))
> plot(forecast(fit11, h=24))
> plot(forecast(fit12, h=24))
```

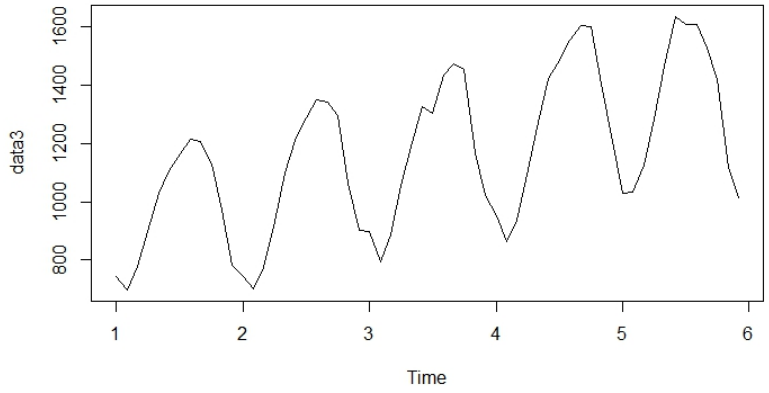




Using dummy variables and apply lm(.) function for multiple linear regression also get the same results as we obtained by tslm(.).

Q. 3)

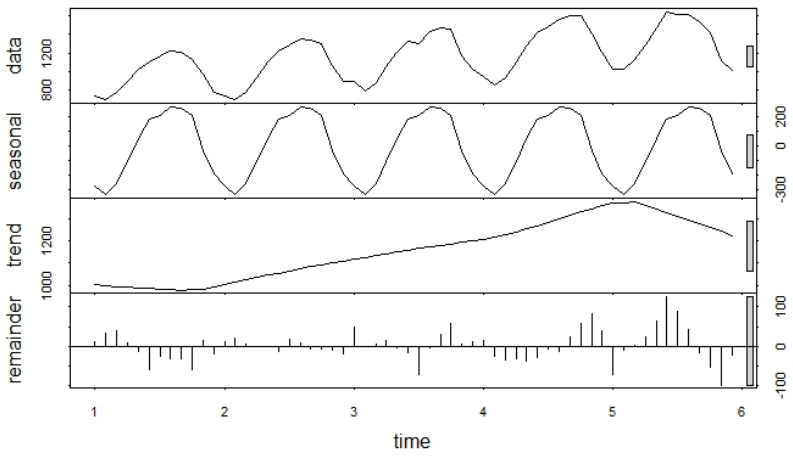
```
> require(fpp)
> data("plastics")
> head(plastics)
[1]  742  697  776  898 1030 1107
> data3 = plastics
> data3
   Jan  Feb  Mar  Apr  May  Jun  Jul  Aug  Sep  Oct  Nov  Dec
1  742  697  776  898 1030 1107 1165 1216 1208 1131  971  783
2  741  700  774  932 1099 1223 1290 1349 1341 1296 1066  901
3  896  793  885 1055 1204 1326 1303 1436 1473 1453 1170 1023
4  951  861  938 1109 1274 1422 1486 1555 1604 1600 1403 1209
5 1030 1032 1126 1285 1468 1637 1611 1608 1528 1420 1119 1013
> par(mfrow=c(1,1))
> plot(data3)
```



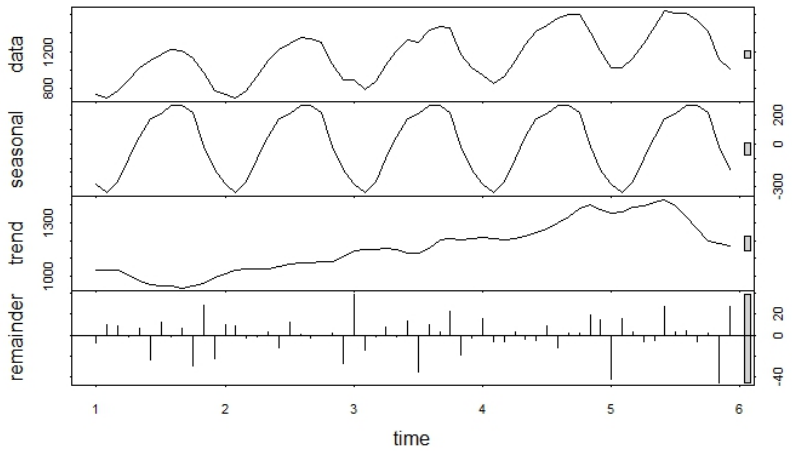

```
> TS3 = ts(data=data3, frequency = 12, start=c(1, 1), end=c(5, 12))
```

As one can see there is a seasonal fluctuations and a trend.

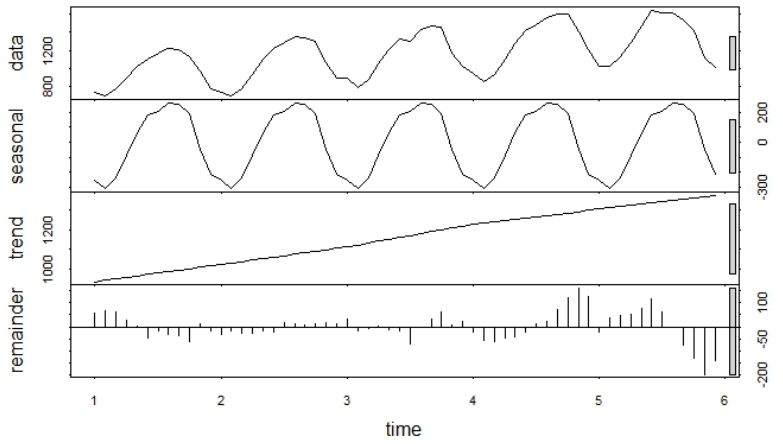
```
> fit31 = stl(TS3 , s.window="periodic")
> plot(fit31)
```



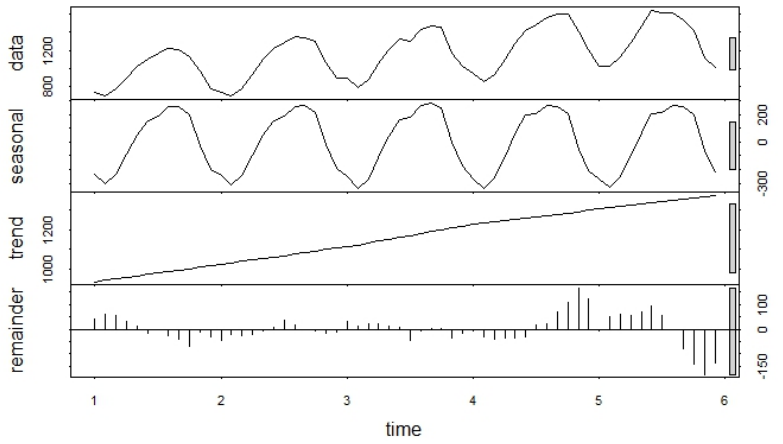

```
> fit32 = stl(TS3 , s.window="periodic", t.window=5)
> plot(fit32)
```




```
> fit33 = stl(TS3 , s.window="periodic", t.window=50)
> plot(fit33)
```

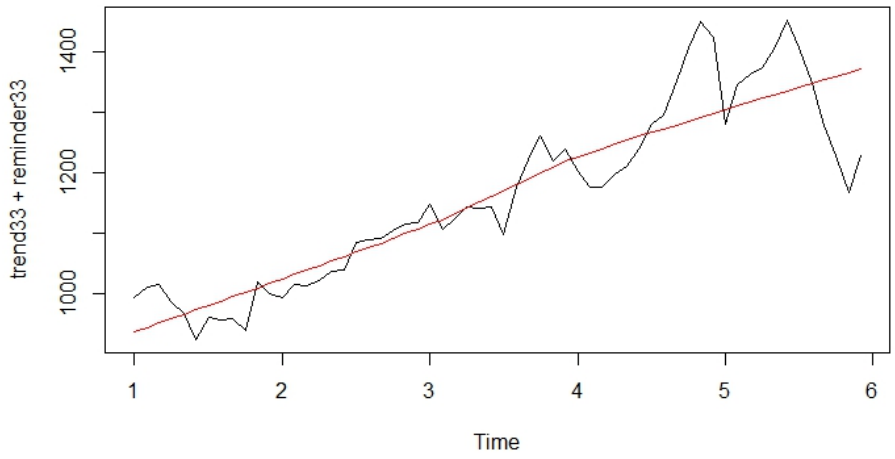



```
> fit34 = stl(TS3 , s.window=5, t.window=50)
> plot(fit34)
```

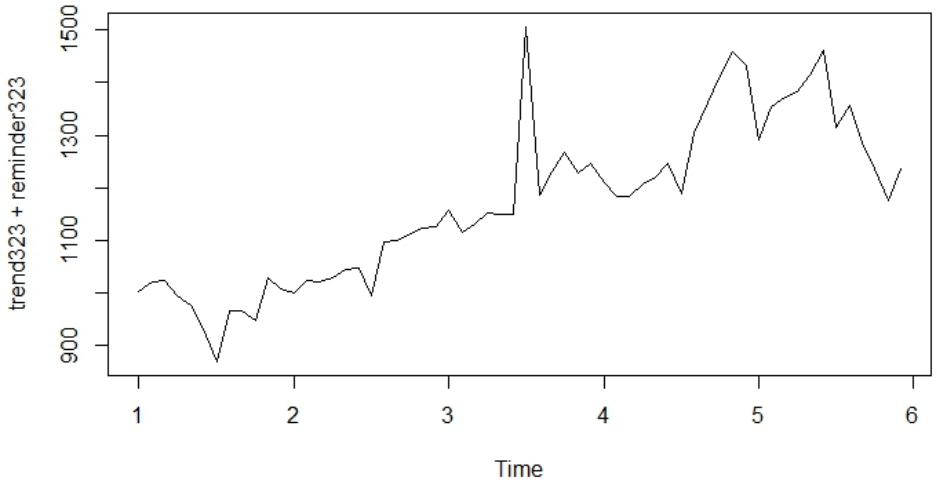


There are changes in the shape of “trend” and “reminder” components of our decompositions, w hen changing the seasonal and trend windows. When we use a small window for trend and seaso nality, we will see more localized estimations on trend and seasonality.

```
> reminder33 <- fit33$time.series[,3]
> trend33 <- fit33$time.series[,2]
> seasonal33 <- fit33$time.series[,1]
> plot(trend33 + reminder33)
> lines(trend33,col="red")
```



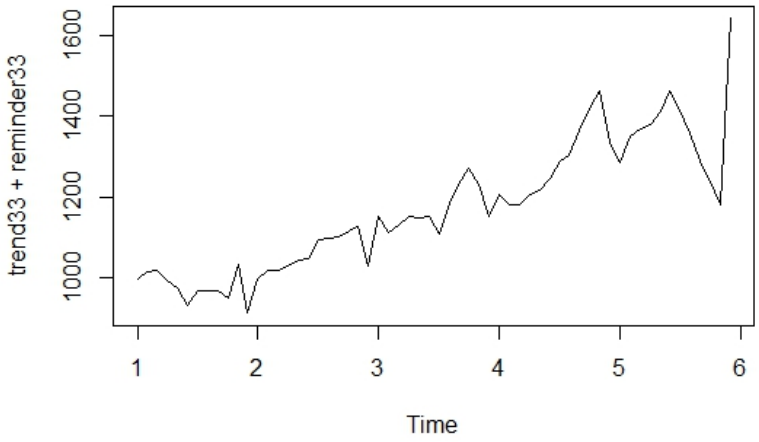

```
> data32= data3
> data32[31]=data32[31]+500
> TS32 = ts(data=data32, frequency = 12, start=c(1, 1), end=c(5, 12))
> fit323 = stl(TS32 , s.window="periodic", t.window=50)
> seasonal323 <- fit323$time.series[,1]
> trend323 <- fit323$time.series[,2]
> reminder323 <- fit323$time.series[,3]
> plot(trend323 + reminder323)
```



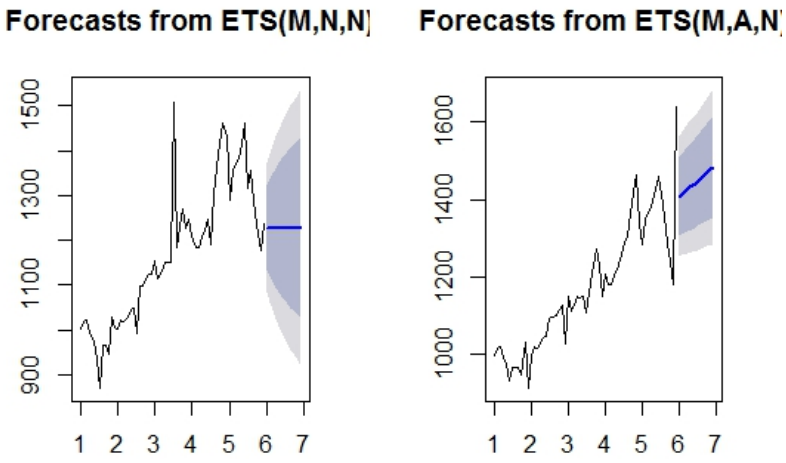

```
> data33=data3
> data33[length(data33)]=data33[length(data33)]+500
> data33
   Jan  Feb  Mar  Apr  May  Jun  Jul  Aug  Sep  Oct  Nov  Dec
1  742  697  776  898 1030 1107 1165 1216 1208 1131  971  783
2  741  700  774  932 1099 1223 1290 1349 1341 1296 1066  901
```

```
3  896  793  885 1055 1204 1326 1303 1436 1473 1453 1170 1023
4  951  861  938 1109 1274 1422 1486 1555 1604 1600 1403 1209
5 1030 1032 1126 1285 1468 1637 1611 1608 1528 1420 1119 1513
```

```
> TS33 = ts(data=data33, frequency = 12, start=c(1, 1), end=c(5, 12))
> fit33 = stl(TS33 , s.window="periodic", t.window=50)
> seasonal33 <- fit33$time.series[,1]
> trend33 <- fit33$time.series[,2]
> reminder33 <- fit33$time.series[,3]

> plot(trend33 + reminder33)
```




```
> par(mfrow=c(1,2))
> plot(forecast(trend323 + reminder323, h=12))
> plot(forecast(trend33 + reminder33, h=12))
```



As one can see, if outlier appears in last observations it will have a much bigger impact on foreca sts than when the outlier appears in the middle of the time series.