STATISTICS 3N03/3J04

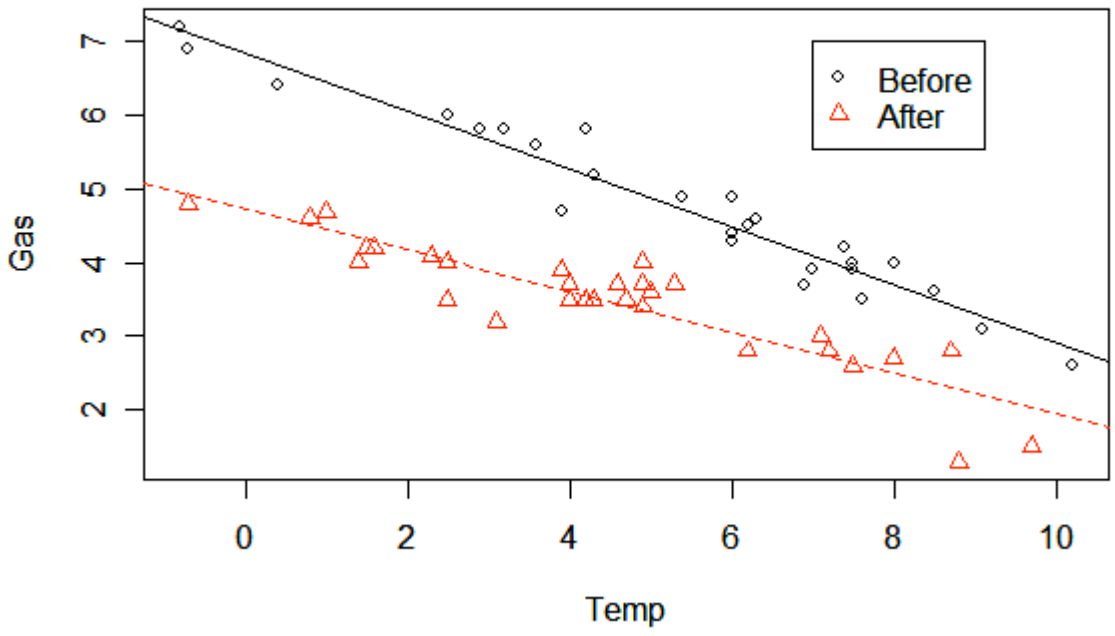
Test 1a – 2006-10-03 – Solutions

Question 1. **[12 marks for appropriate conclusions supported by graphs.]**

The gas consumption required to heat Mr Whiteside's house appears to be linearly related to temperature, with negative slope. After the insulation was installed, the slope of this relationship was reduced. The insulation saved about 2000 ft^3 of gas per week when the mean weekly temperature was 0□ C, compared to about 1000 ft^3 per week at 10□ C.

If you inspect the whiteside data frame, the rows have evidently been sorted by Temp within Insul. This means that the time order of the observations has been lost so we can't analyze them as a time series.

```
> library(MASS)
> names(whiteside)
[1] "Insul" "Temp"  "Gas"
> levels(whiteside$Insul)
[1] "Before" "After"
> plot(Gas~Temp, whiteside, pch=as.numeric(Insul), col=as.numeric(Insul))
> abline(lm(Gas~Temp, whiteside[whiteside$Insul=="Before",]), col=1, lty=1)
> abline(lm(Gas~Temp, whiteside[whiteside$Insul=="After",]), col=2, lty=2)
> legend(7, 7, levels(whiteside$Insul), col=1:2, pch=1:2)
```



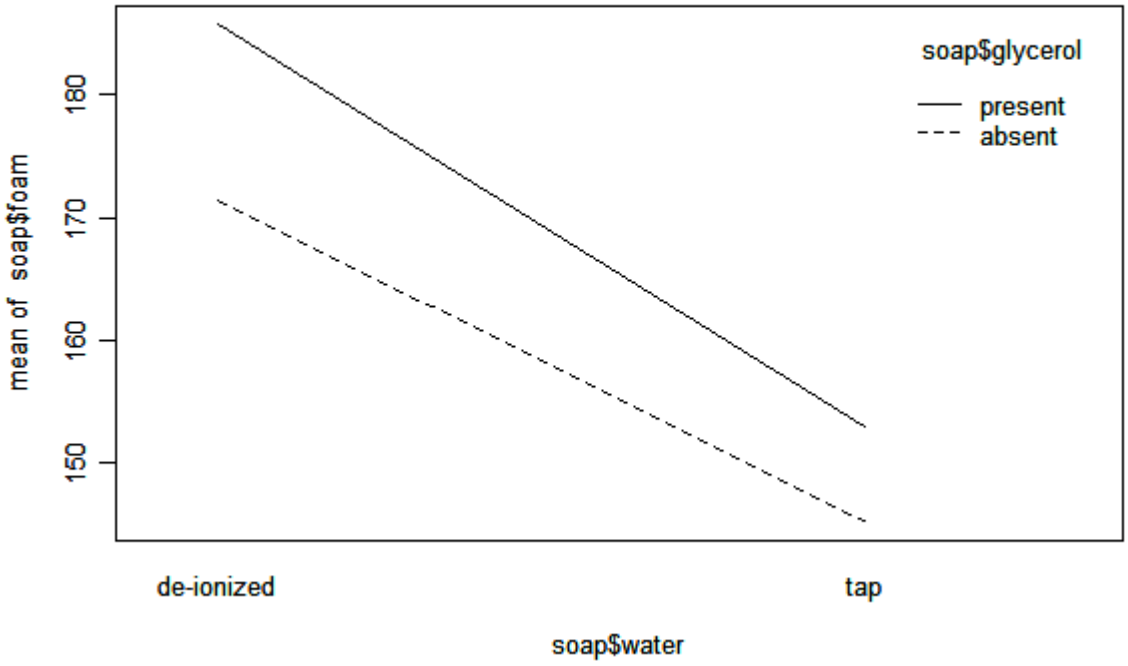
Question 2. **[12 marks for appropriate conclusions supported by graphs.]**

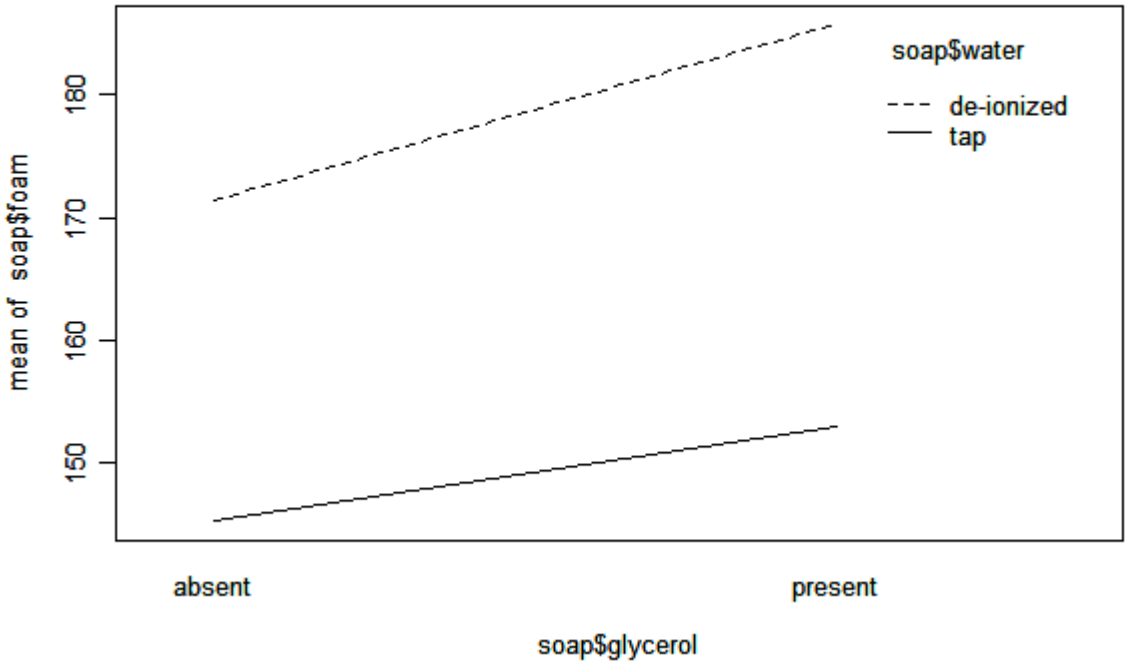
From either interaction plot, we can see that glycerol increases the amount of foam, and de-ionized water gives more foam than tap water, and the two factors do not appear to interact (the lines are reasonably parallel).

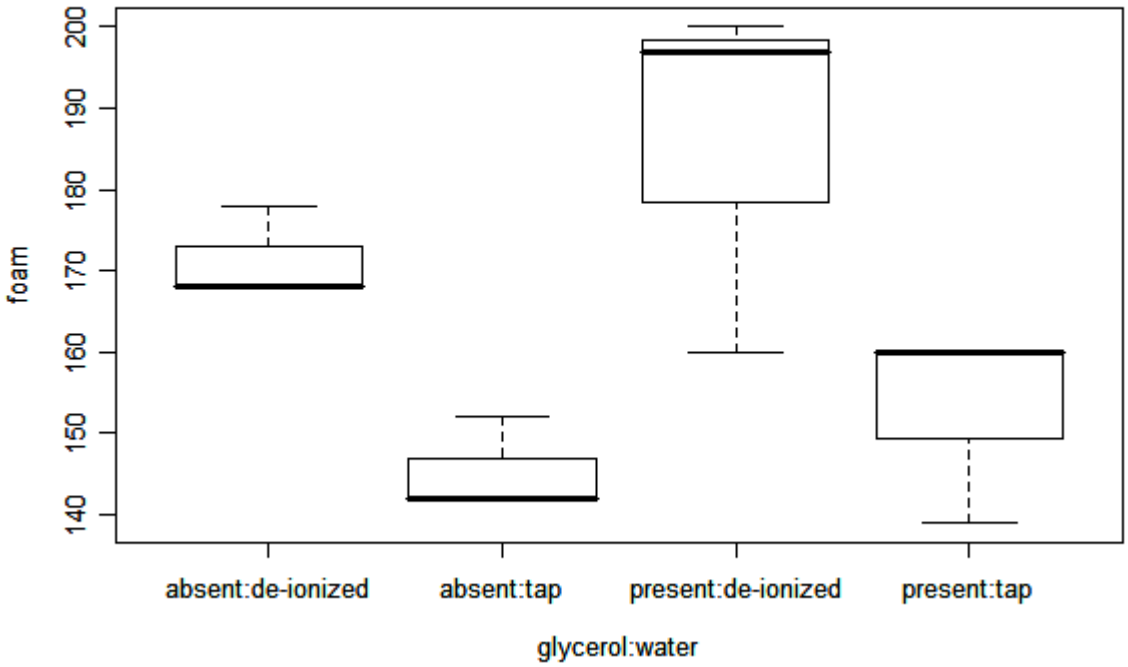
The box plots lead to the same conclusion, but also show the variation; it is interesting that without glycerol the variation is above the median, with glycerol there is more variation and it is below the median.

Box plots of each factor separately would also be interesting, but because both factors have strong effects they will not show the differences so clearly.

```
> soap <- data.frame(foam=c(168, 178, 168, 152, 142, 142, 160, 197, 200, 139, 160, 160),
glycerol=factor(rep(c("absent","present"),c(6,6))), water=factor(rep(rep(c("de-
ionized","tap"),c(3,3)),2)))
> soap
   foam glycerol      water
1   168   absent de-ionized
2   178   absent de-ionized
3   168   absent de-ionized
4   152   absent        tap
5   142   absent        tap
6   142   absent        tap
7   160  present de-ionized
8   197  present de-ionized
9   200  present de-ionized
10  139  present        tap
11  160  present        tap
12  160  present        tap
> par(cex=.8)
> interaction.plot(soap$water,soap$glycerol,soap$foam)
> interaction.plot(soap$glycerol,soap$water,soap$foam)
> boxplot(split(soap$foam, soap$glycerol:soap$water), xlab="glycerol:water", ylab="foam")
```


soap$glycerol
present
absent
mean of soap$foam
150
160
170
180
de-ionized
tap
soap$water

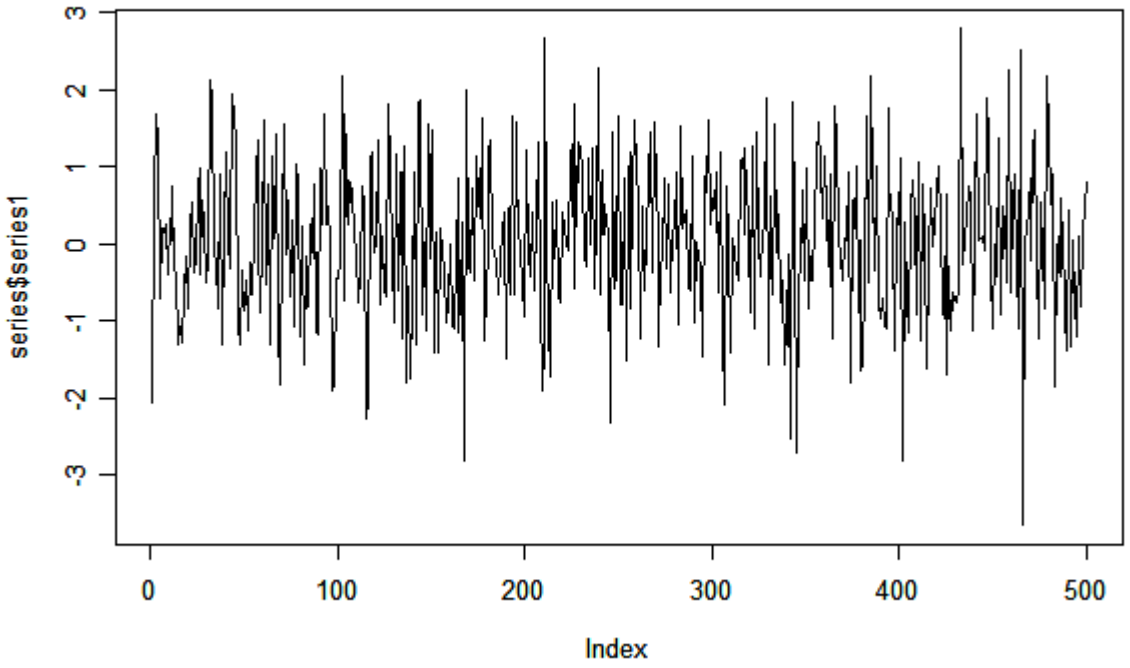

soap$water
de-ionized
tap
mean of soap$foam
150
160
170
180
absent
present
soap$glycerol

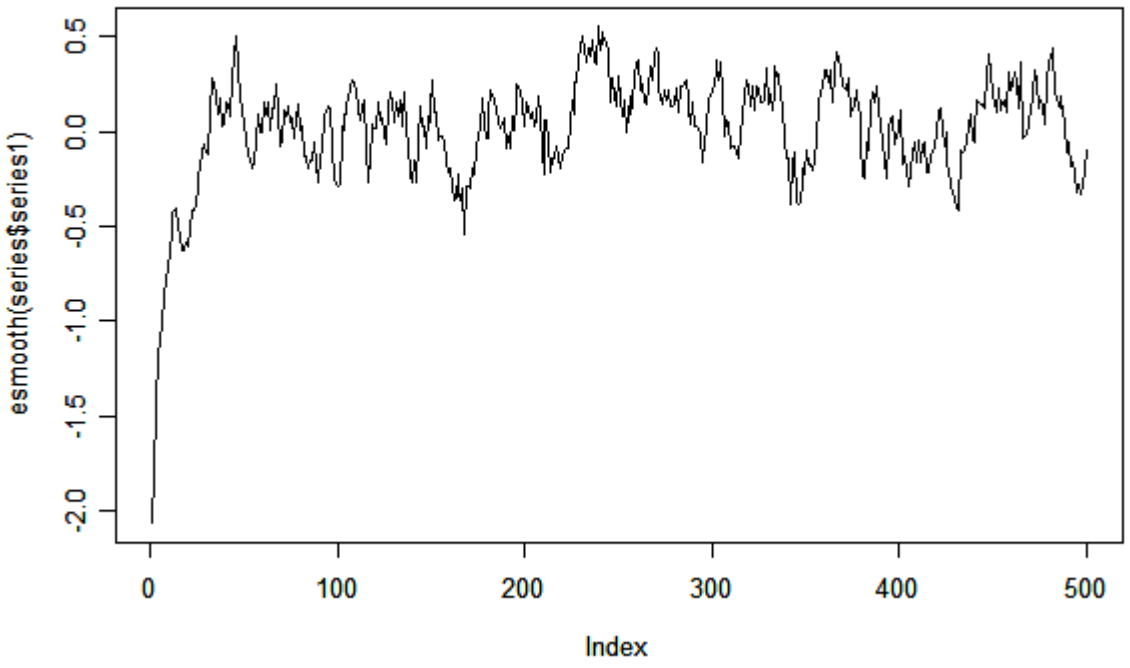

foam
200
190
180
170
160
150
140
absent:de-ionized
absent:tap
present:de-ionized
present:tap
glycerol:water

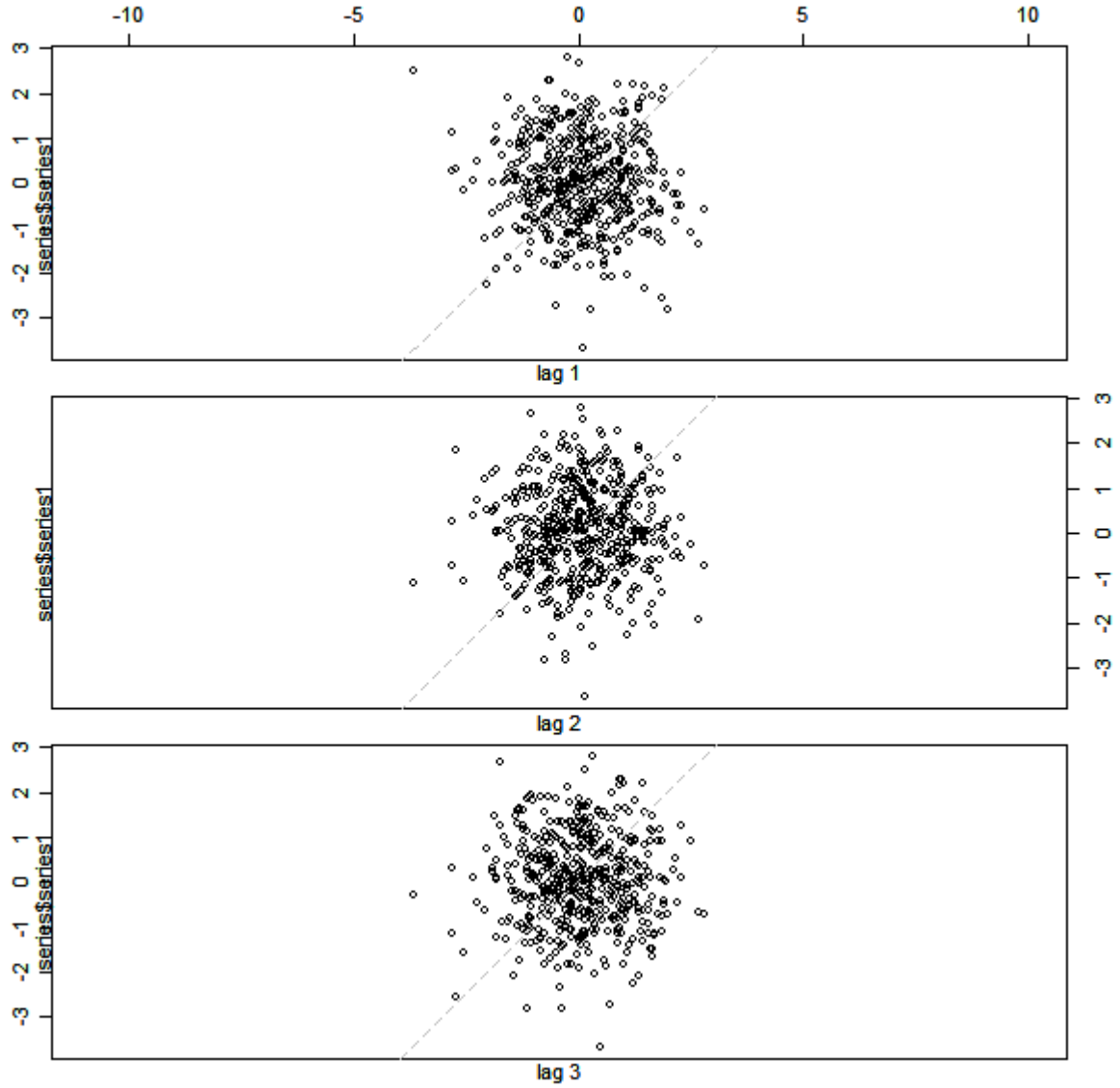
Question 3. **[12 marks for appropriate conclusions supported by graphs.]**

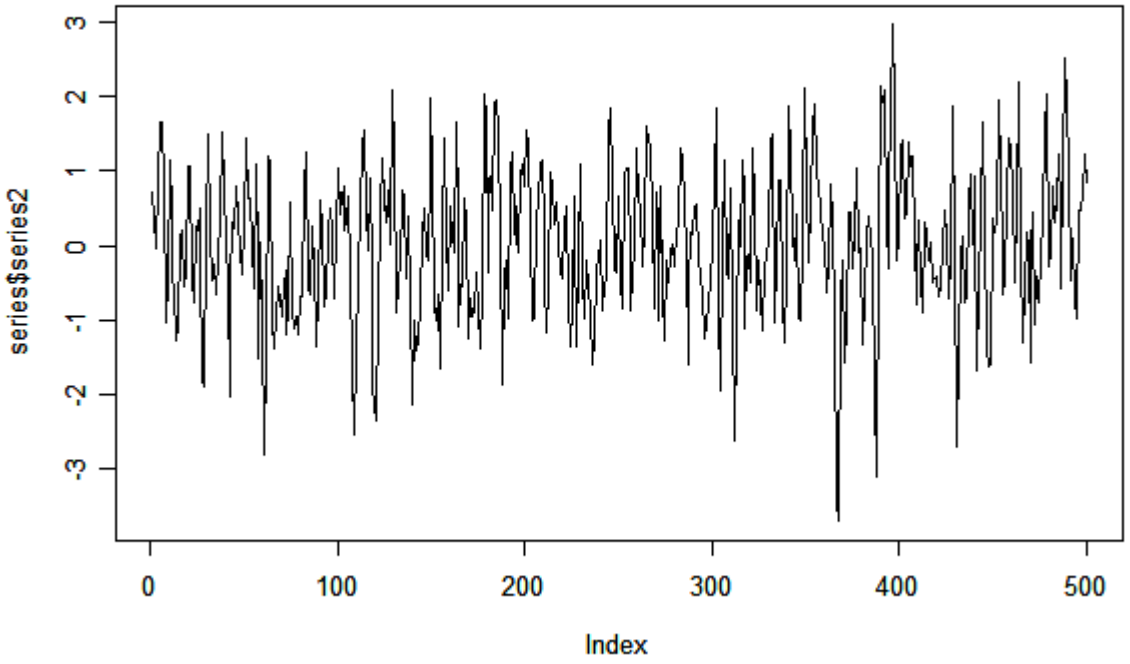
The lag plots clearly show that the observations in Series 1 are independent, observations in Series 2 are autocorrelated with lag 1, observations in Series 3 are autocorrelated with lag 2. The time sequence plots confirm this but the effect is less obvious; Series 1 oscillates on a scale of 1 time step, Series 2 oscillations are about 2 time steps wide, Series 3 oscillations are about 3 time steps wide. The plot for Series 1 is the most dense of the three, Series 3 the least dense. Smoothing the series does not seem to be helpful here, the smoothed series all look about the same after the effect of the first observation has worn off.

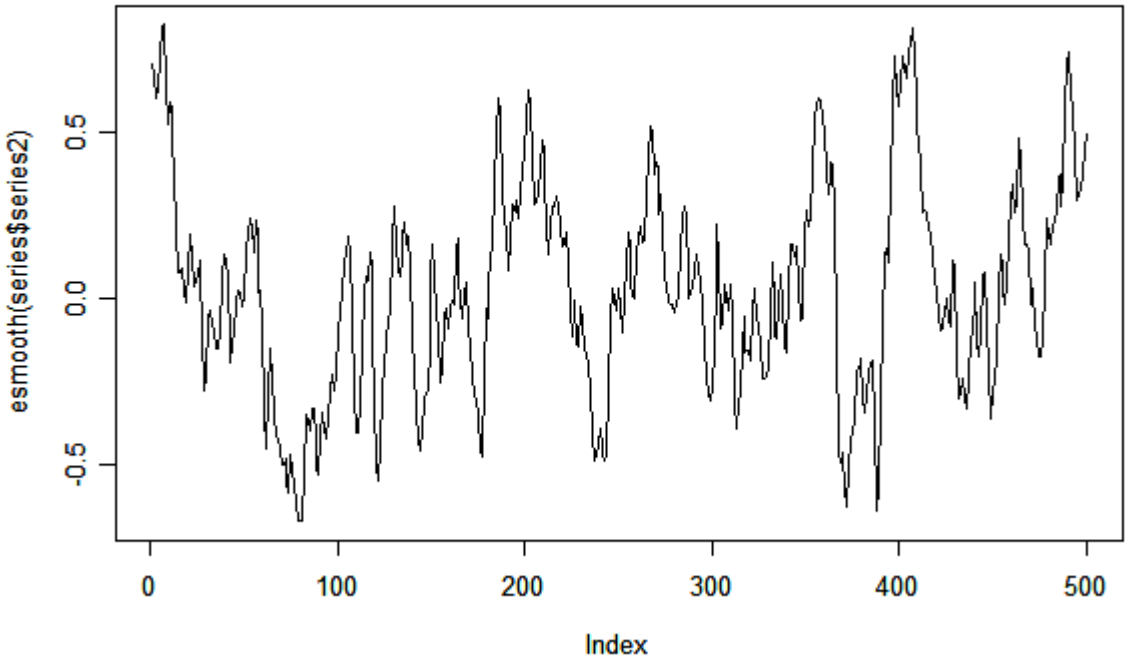
```
> smooth <- function (series, alpha = 0.1)
{
    esseries = series
    for (t in 2:length(series)) esseries[t] <- ifelse(is.na(series[t]),
        esseries[t - 1], alpha * series[t] + (1 - alpha) * esseries[t -
            1])
    esseries
}
> series <- read.table("series.txt")
> names(series)
[1] "series1" "series2" "series3"
> dim(series)
[1] 500   3
> plot(series$series1, type="l")
> plot(esmooth(series$series1), type="l")
> lag.plot(series$series1, 3)
> plot(series$series2, type="l")
> plot(esmooth(series$series2), type="l")
> lag.plot(series$series2, 3)
> plot(series$series3, type="l")
> plot(esmooth(series$series3), type="l")
> lag.plot(series$series3, 3)
```

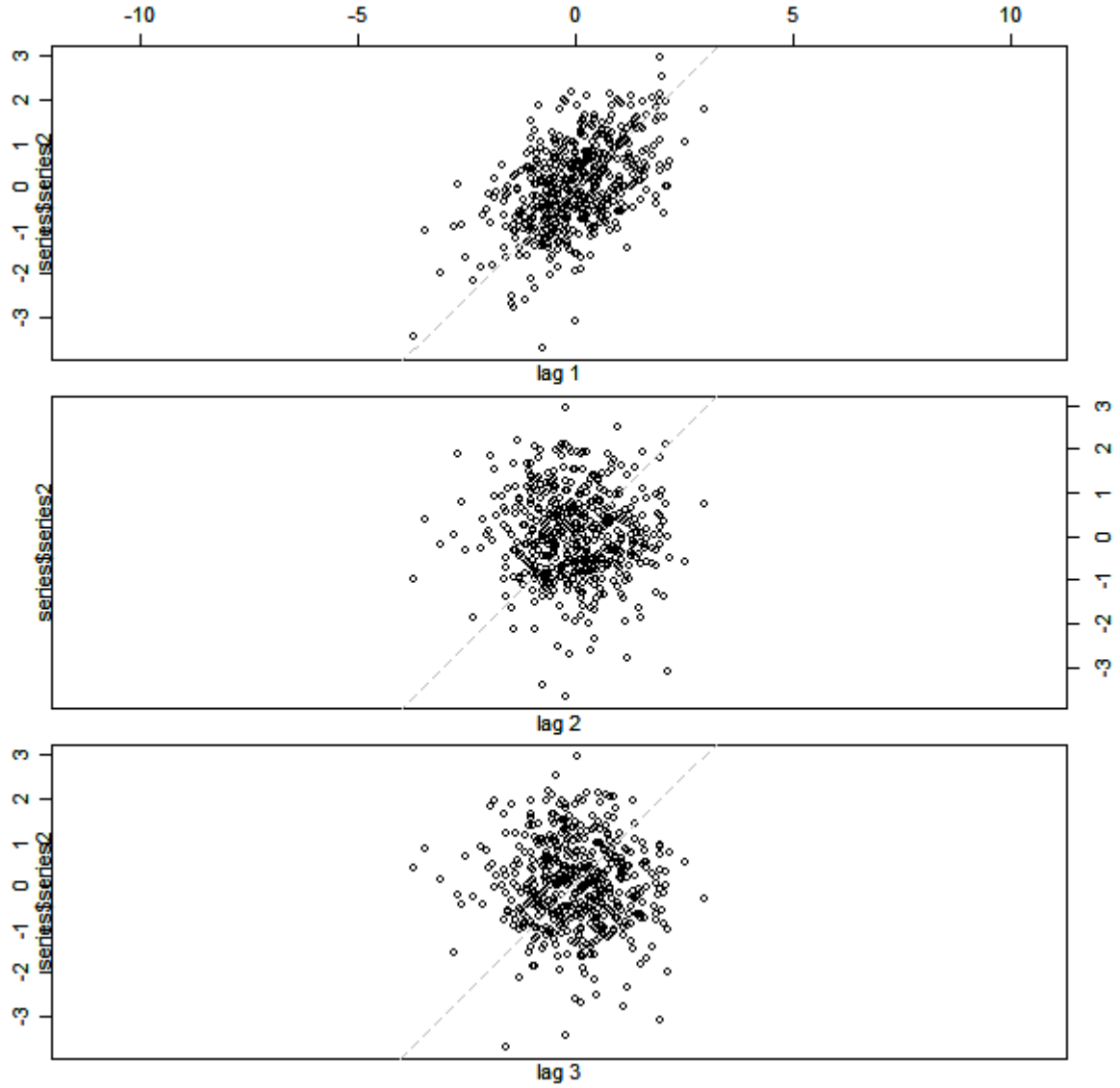

series$series1
3
2
1
0
-1
-2
-3
0
100
200
300
400
500
Index

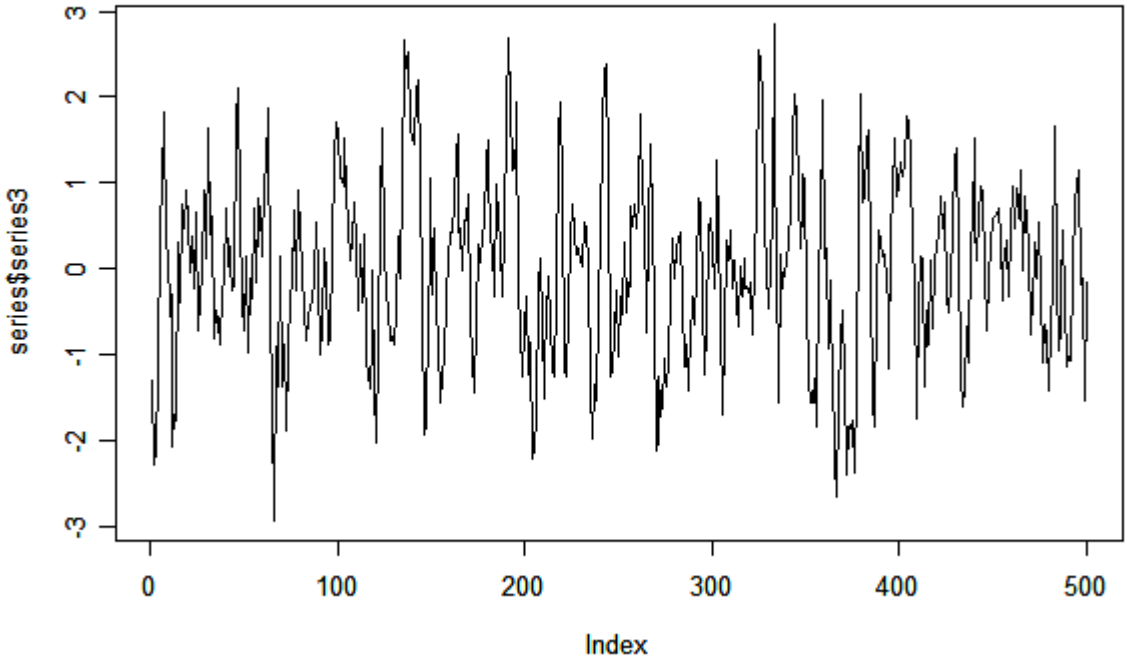

esmooth(series$series1)
0.5
0.0
-0.5
-1.0
-1.5
-2.0
0
100
200
300
400
500
Index

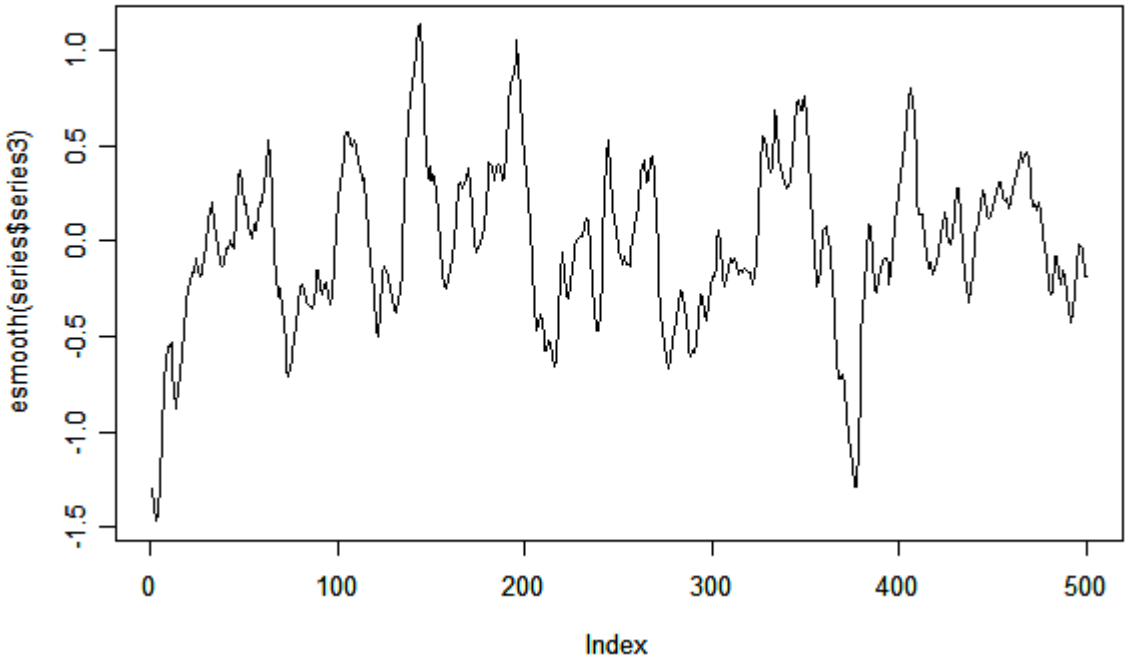

-10
-5
0
5
10
lseries$series1
lag 1
series$series1
lag 2
lseries$series1
lag 3

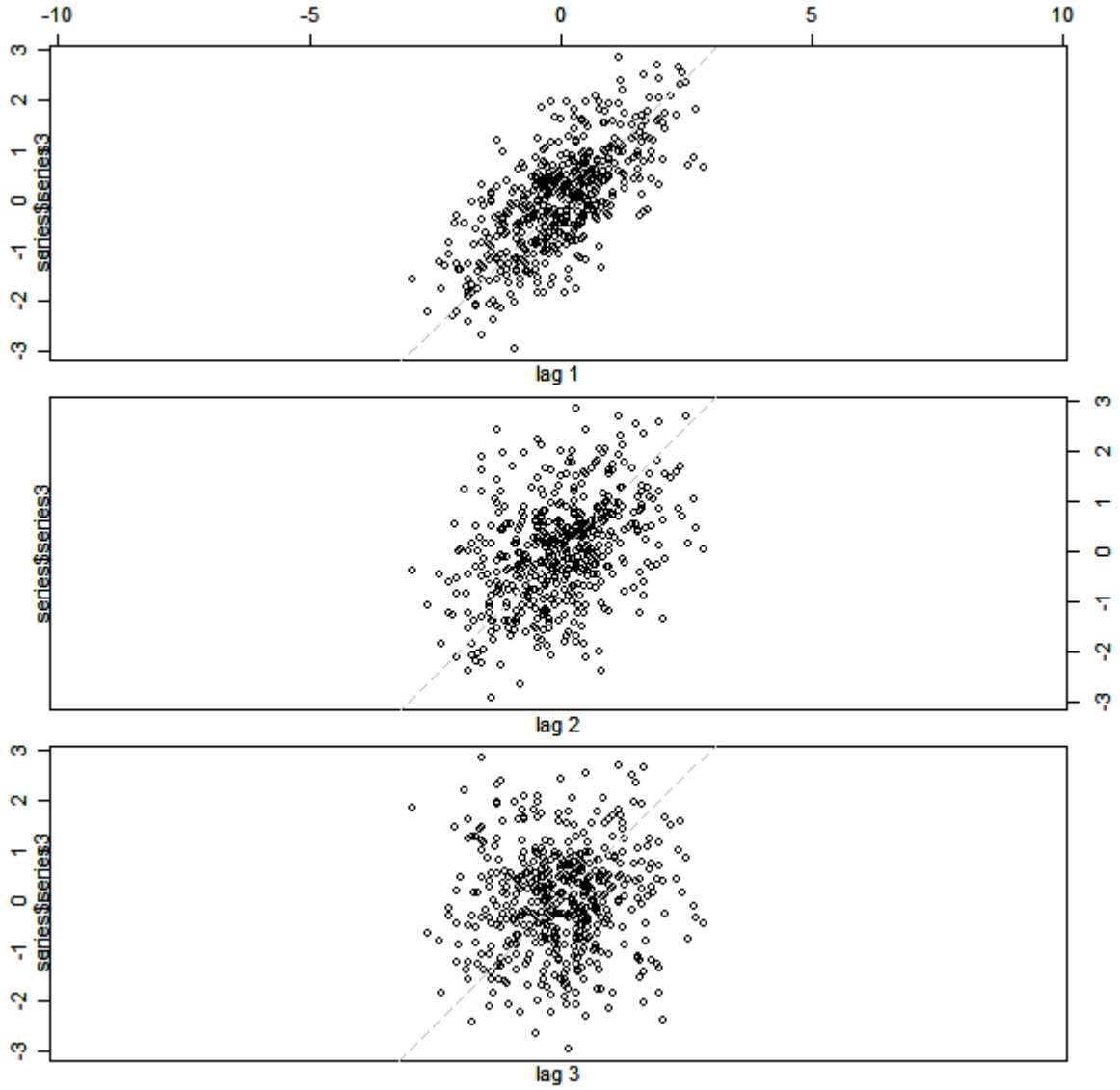

series$series2
3
2
1
0
-1
-2
-3
0
100
200
300
400
500
Index


esmooth(series$series2)
0.5
0.0
-0.5
0
100
200
300
400
500
Index


-10
-5
0
5
10
series$series2
lag 1
lag 2
lag 3


series$series3
3
2
1
0
-1
-2
-3
0
100
200
300
400
500
Index


esmooth(series$series3)
1.0
0.5
0.0
-0.5
-1.0
-1.5
0
100
200
300
400
500
Index



BONUS MARKS: [Up to 9 marks for the quality of writing and presentation]

FULL MARKS = 45