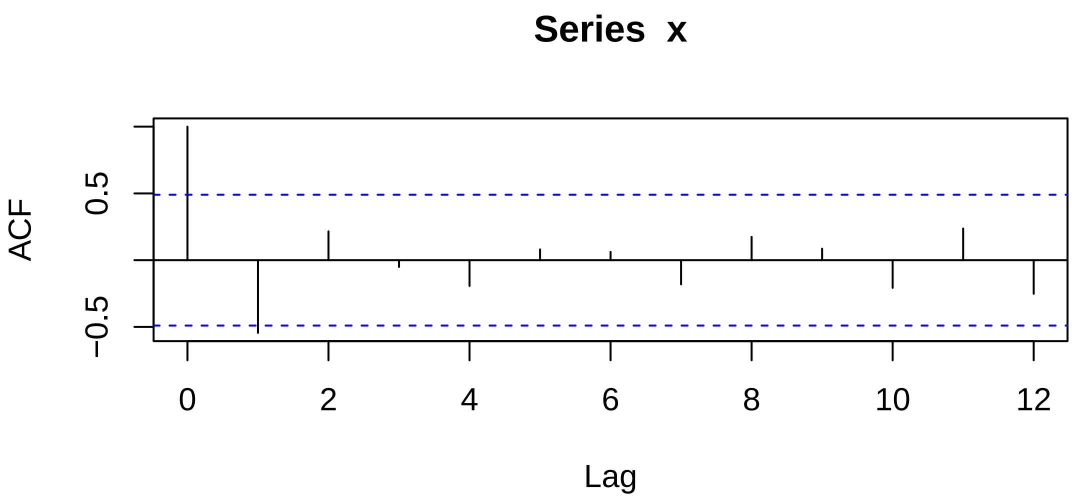
lab-2-stat3925

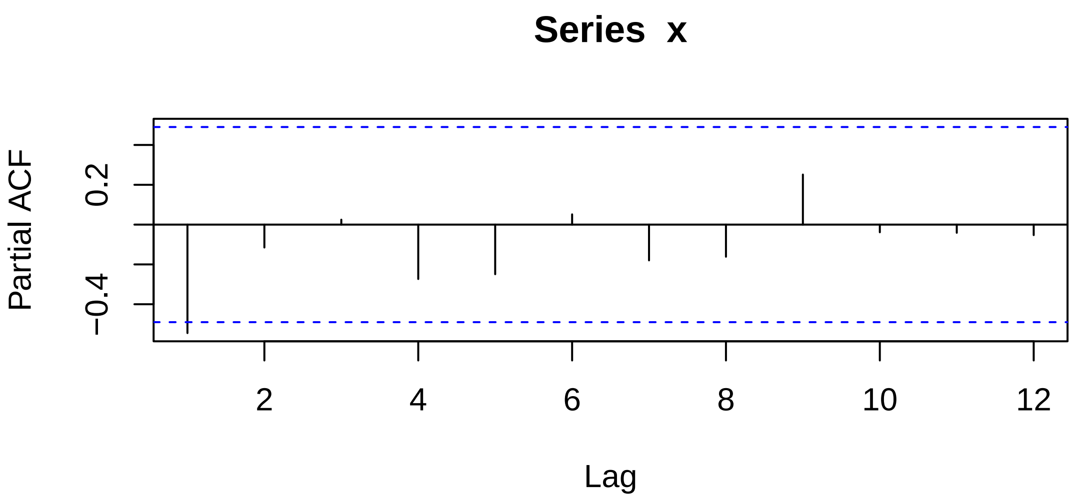
Mason

27/02/2022

Question 1

```
par(mfrow = c(2, 1))
x = c(1.6, 0.8, 1.2, 0.5, 0.9, 1.1, 1.1, 0.6, 1.5, 0.8, 0.9, 1.2, 0.5, 1.3, 0.8,
    1)
acf(x)
pacf(x)
```



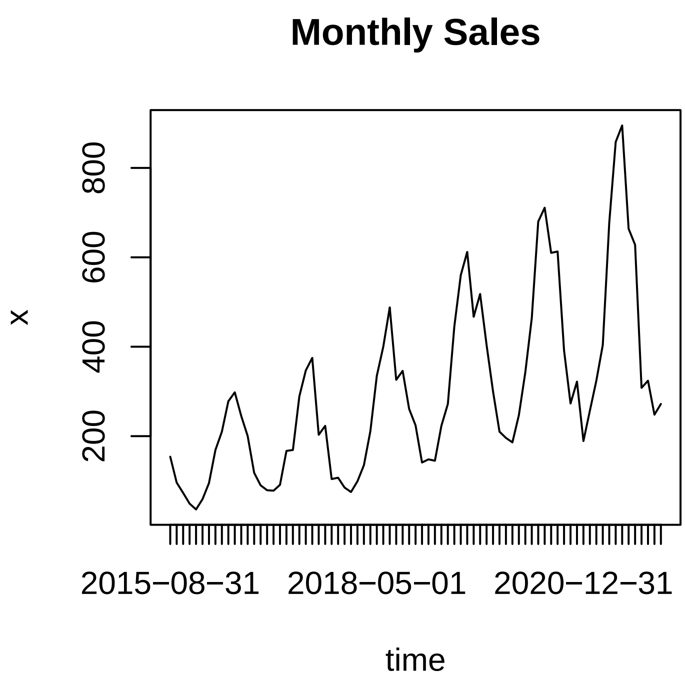


The data doesn't seem to be stationary as it doesn't quickly decay. There seems to be a seaonal component present.

Question 2

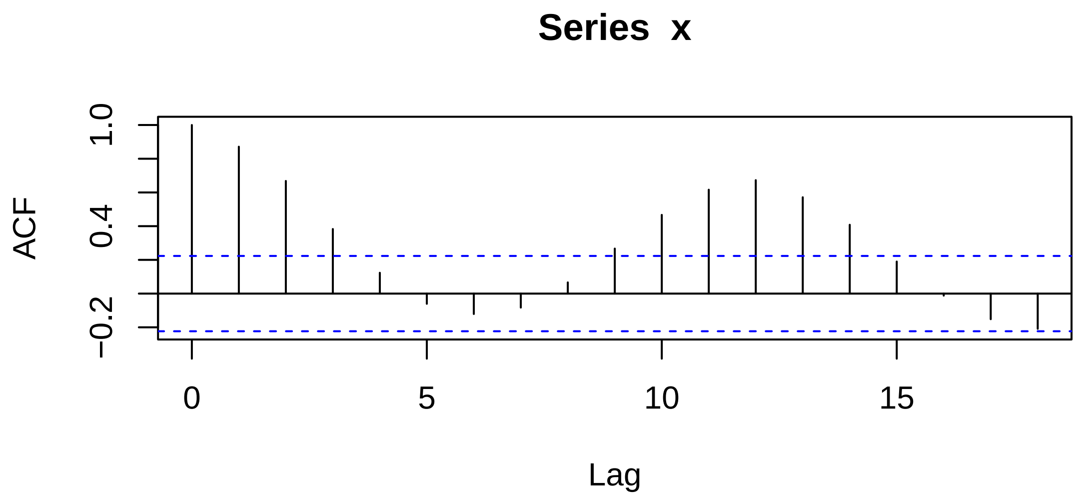
1. In the following plot we plot the time series object, representing the sales of an item of a company in successive months from August 2015.
2. We comment that the data doesn't have constant mean and doesn't have constant variance. There seems to be a seasonal trend with $p = 12$ (that is every year). There seems to be a linear trend.
3. In the following plot we have properly labelled the axes and given the time series a title.

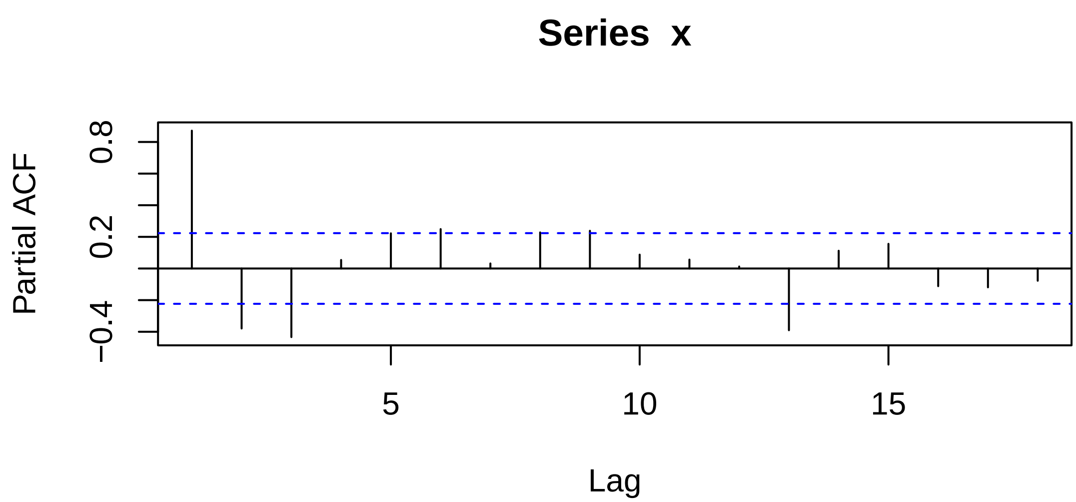
```
x = c(154, 96, 73, 49, 36, 59, 95, 169, 210, 278, 298, 245, 200, 118, 90, 79, 78,
    91, 167, 169, 289, 347, 375, 203, 223, 104, 107, 85, 75, 99, 135, 211, 335, 400,
    488, 326, 346, 261, 224, 141, 148, 145, 223, 272, 445, 560, 612, 467, 518, 404,
    300, 210, 196, 186, 247, 343, 464, 680, 711, 610, 613, 392, 273, 322, 189, 257,
    324, 404, 677, 858, 895, 664, 628, 308, 324, 248, 272)

x1 = ts(x, start = c(2015, 8), frequency = 12)  # using August 2015 as the start date.
tsp = attributes(x1)$tsp
dates = seq(as.Date("2015-08-31"), by = "month", along = x1)
plot(x1, xaxt = "n", main = "Monthly Sales", ylab = "x", xlab = "time")
axis(1, at = seq(tsp[1], tsp[2], along = x1), labels = format(dates))
```



4. We now plot the acf and the pacf of the series in x in the following two plots. There is a **very clear** seasonal trend which is decaying. There are statistically significant correlation values at $k = 1, 2, 3, 9, 10, 11, 12, 13, 14$. So we would say that this is also not a random time series. There is a tend (seasonal and possibly a linear trend present)

```
par(mfrow = c(2, 1))
acf(x)
pacf(x)
```

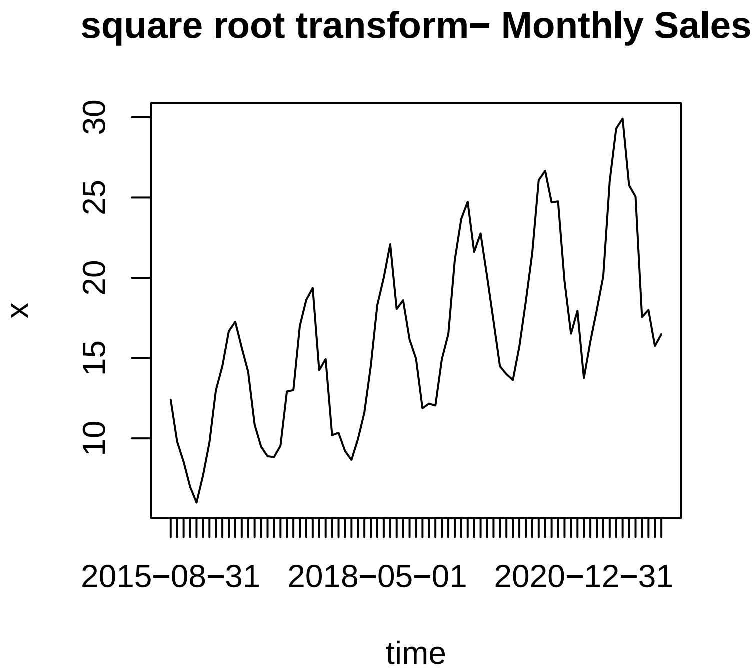




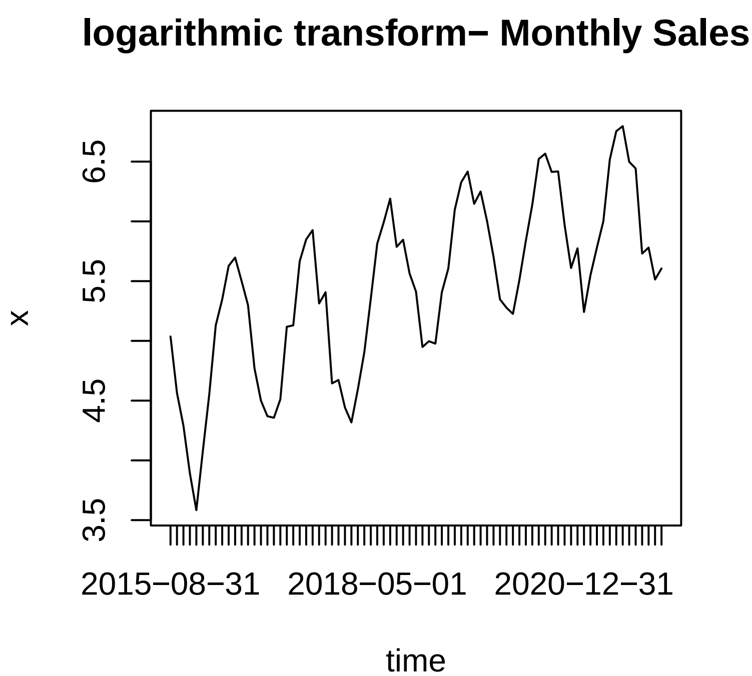
5. I think that the previous series need to be transformed for further analysis because it is a non-stationary time series. That is, it **doesn't have constant mean** and **doesn't have constant second moment** (inferred because the variance isn't constant.)

6. We now use the the square root transformation and the logarithmic transformation of the data. We plot both the time series

```
x2 = ts(sqrt(x), start = c(2015, 8), frequency = 12)  # using August 2015 as the start date.
tsp = attributes(x2)$tsp
dates = seq(as.Date("2015-08-31"), by = "month", along = x2)
plot(x2, xaxt = "n", main = "square root transform- Monthly Sales", ylab = "x", xlab = "time")
axis(1, at = seq(tsp[1], tsp[2], along = x2), labels = format(dates))
```




```r
x3 = ts(log(x), start = c(2015, 8), frequency = 12)  # using August 2015 as the start date.
tsp = attributes(x3)$tsp
dates = seq(as.Date("2015-08-31"), by = "month", along = x3)
plot(x3, xaxt = "n", main = "logarithmic transform- Monthly Sales", ylab = "x", xlab = "time")
axis(1, at = seq(tsp[1], tsp[2], along = x3), labels = format(dates))
```

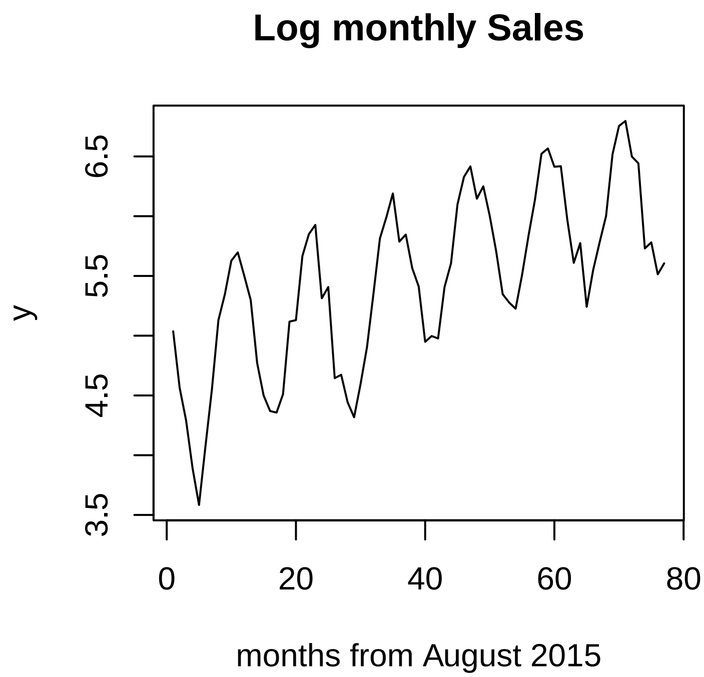


Question 3

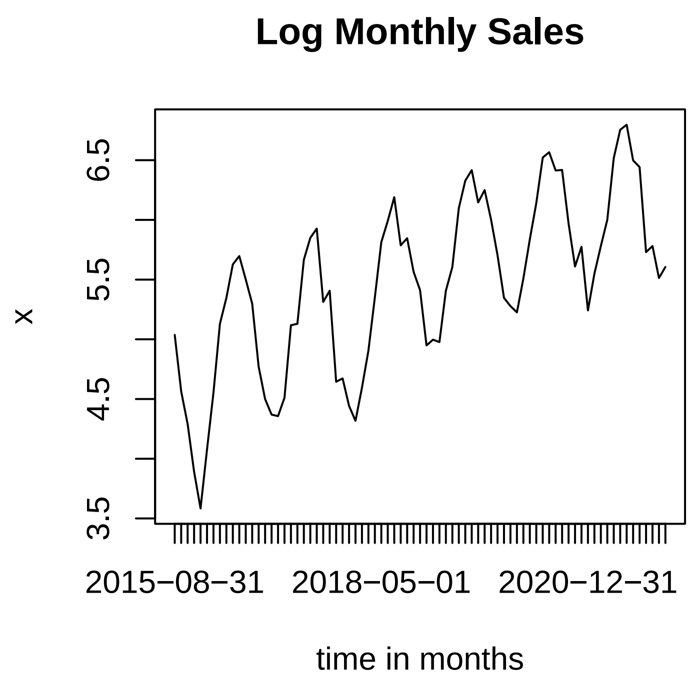
1. Based on the plots in Q2 (vii) we take the series with the logarithmic transformation for further analysis. The reason for this is because with the square root transformation, the variance looks like its still increasing, whereas for the logarithmic transformation the variance looks constant

2. We now plot $y = \log(x)$ and $y_1 = \log(x_1)$. They are the same graph.

```
y = log(x)
plot(y, type = "l", lty = 1, xlab = "months from August 2015")
title("Log monthly Sales")
```

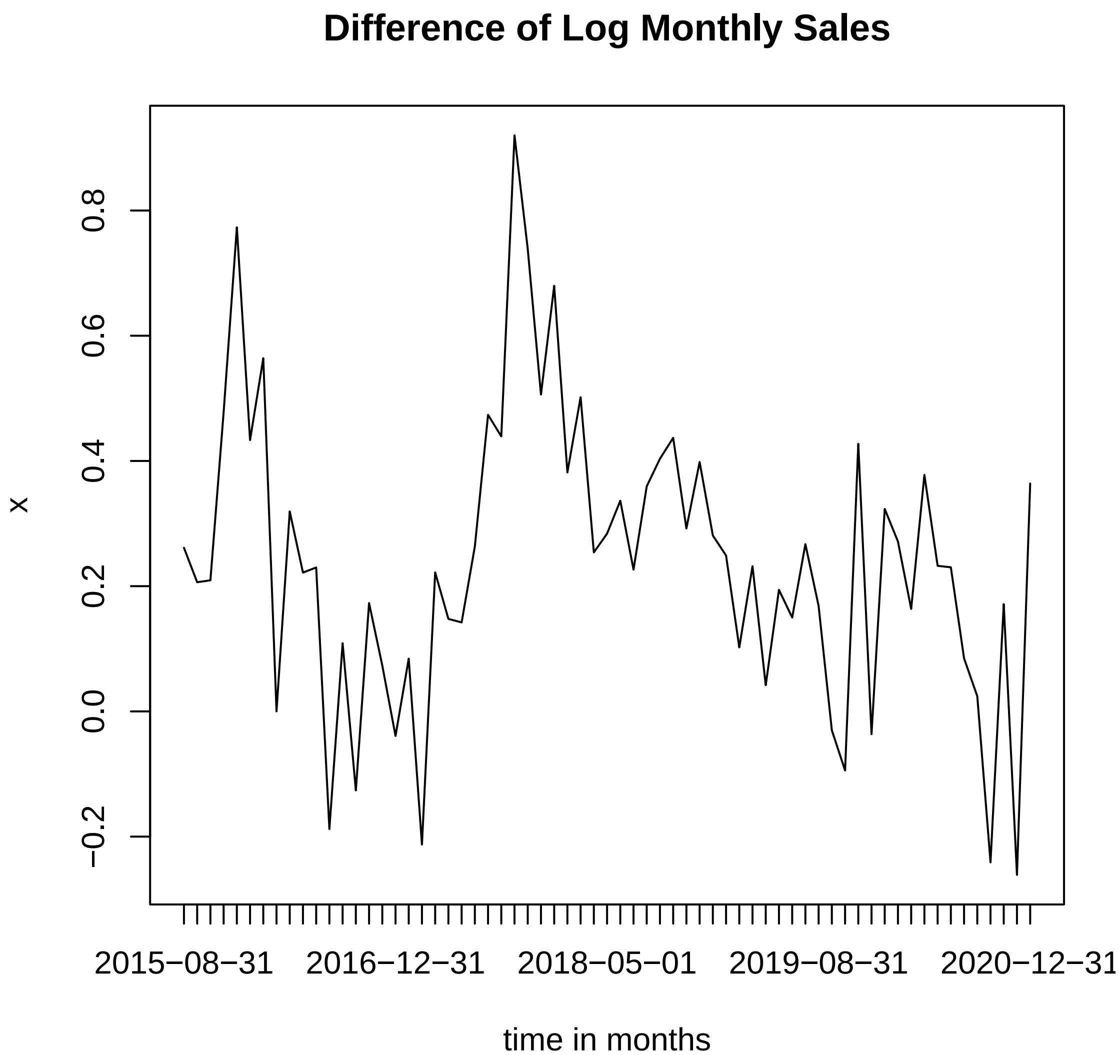



```
y1 = log(x1)
tsp = attributes(y1)$tsp
dates = seq(as.Date("2015-08-31"), by = "month", along = y1)
plot(y1, xaxt = "n", main = "Log Monthly Sales", ylab = "x", xlab = "time in months")
axis(1, at = seq(tsp[1], tsp[2], along = y1), labels = format(dates))
```



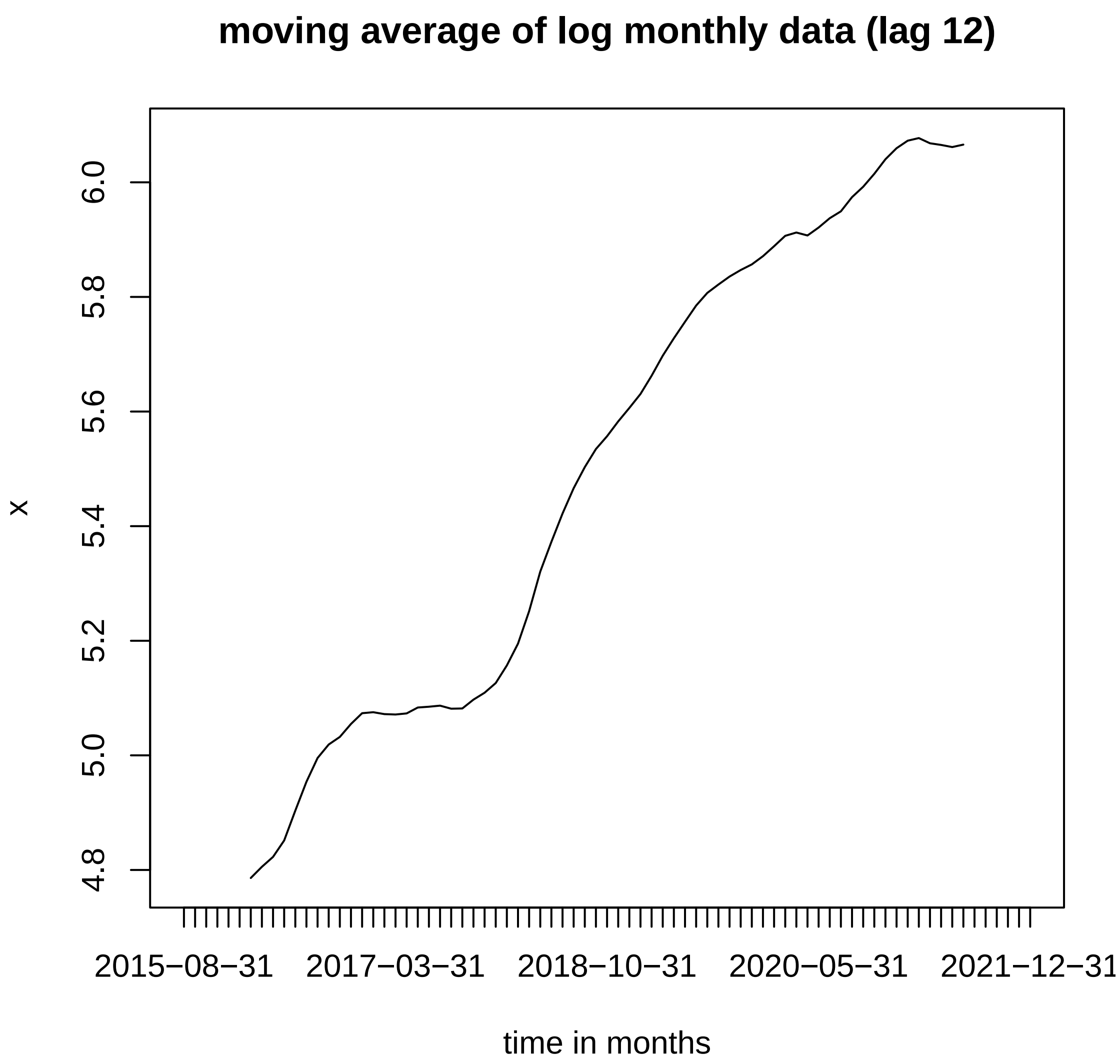
3. We remove the seasonality from the time series in y by a suitable differencing technique. Since the period is $p = 12$ we use lag-12 differencing. We then store this in y_2 and plot it.

```
y2 = diff(y1, lag = 12)
tsp = attributes(y2)$tsp
dates = seq(as.Date("2015-08-31"), by = "month", along = y2)
plot(y2, xaxt = "n", main = "Difference of Log Monthly Sales", ylab = "x", xlab = "time in months")
axis(1, at = seq(tsp[1], tsp[2], along = y2), labels = format(dates))
```



4. A suitable moving average in order to smooth the series y1 is a moving average of size $m = 12$ which is the period.

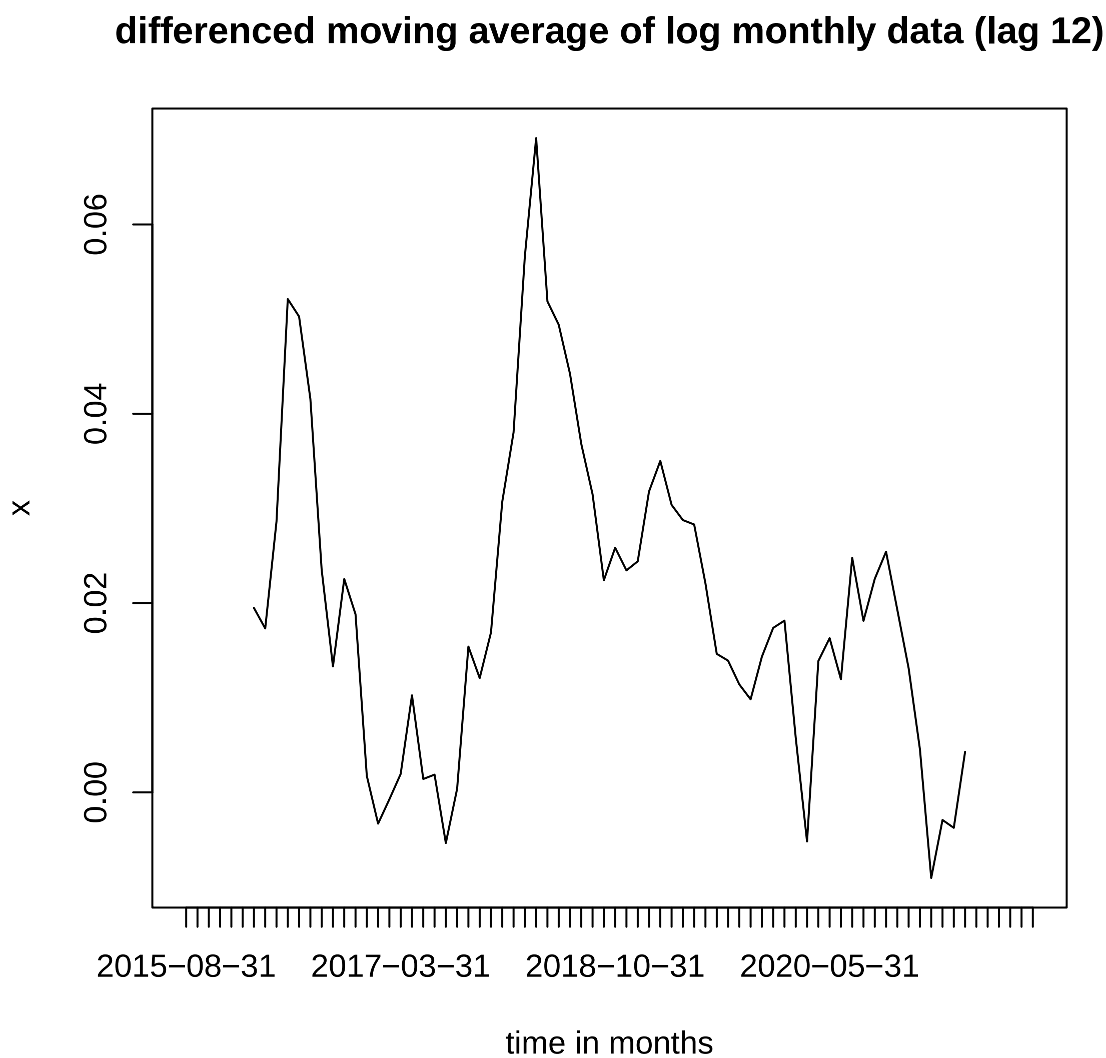
```
m = 12
y3 = filter(y1, c(1/(2 * m), rep(1/m, m - 1), 1/(2 * m)))
tsp = attributes(y3)$tsp
dates = seq(as.Date("2015-08-31"), by = "month", along = y3)
plot(y3, xaxt = "n", main = "moving average of log monthly data (lag 12)", ylab = "x",
    xlab = "time in months")
axis(1, at = seq(tsp[1], tsp[2], along = y3), labels = format(dates))
```



- When we look at the the plot of y_2 we see that this doesn't contain the seasonal or linear trend anymore. The lag-12 differencing removed both the seasonal and linear trend.
- When we look at the plot of y_3 we see that this still contains a linear trend but no seasonal trend anymore. The moving average of size 12 removed the seasonal trend.

5. We can use a lag-1 differencing on y_3 to remove the linear trend.

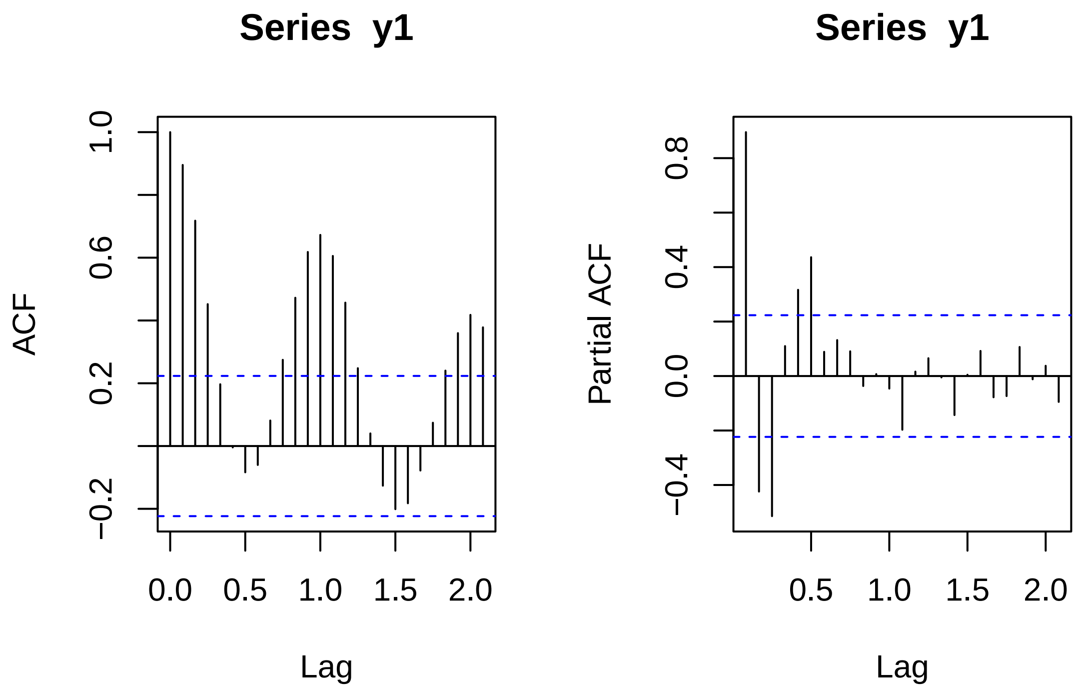
```
y4 = diff(y3, lag = 1)
tsp = attributes(y4)$tsp
dates = seq(as.Date("2015-08-31"), by = "month", along = y4)
plot(y4, xaxt = "n", main = "differenced moving average of log monthly data (lag 12)",
    ylab = "x", xlab = "time in months")
axis(1, at = seq(tsp[1], tsp[2], along = y4), labels = format(dates))
```



- comparing y_4 to x in Q2(iii) we see that this time series has relatively constant variance and doesn't have a linear and seasonal trend.
- comparing y_4 to y in Q3(ii) we see that the variance of y_4 is smaller than that of y. y_4 doesn't have the linear trend whereas y still has a linear trend. They both still have fairly constant variance.
- comparing y_4 to y_2 in Q3(iii) we see that y_4 looks a lot like y_2 but the variance is **a lot smaller**.

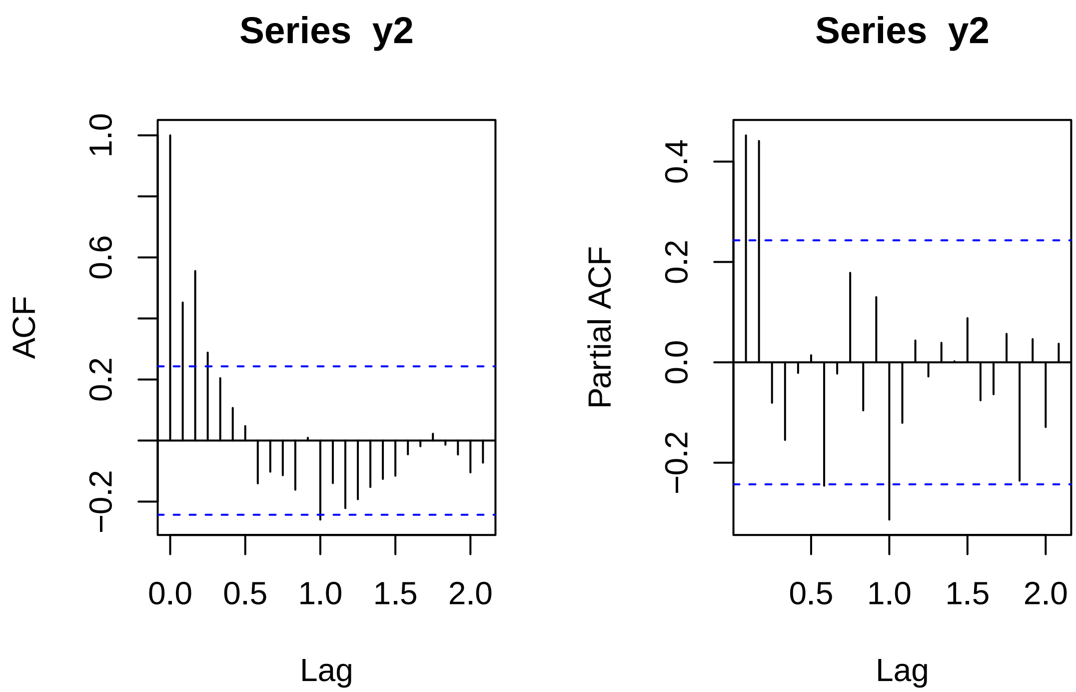
6.

```
par(mfcol = c(1, 2))
y5 = y1 - y3
acf(y1, lag.max = 25)
pacf(y1, lag.max = 25)
```



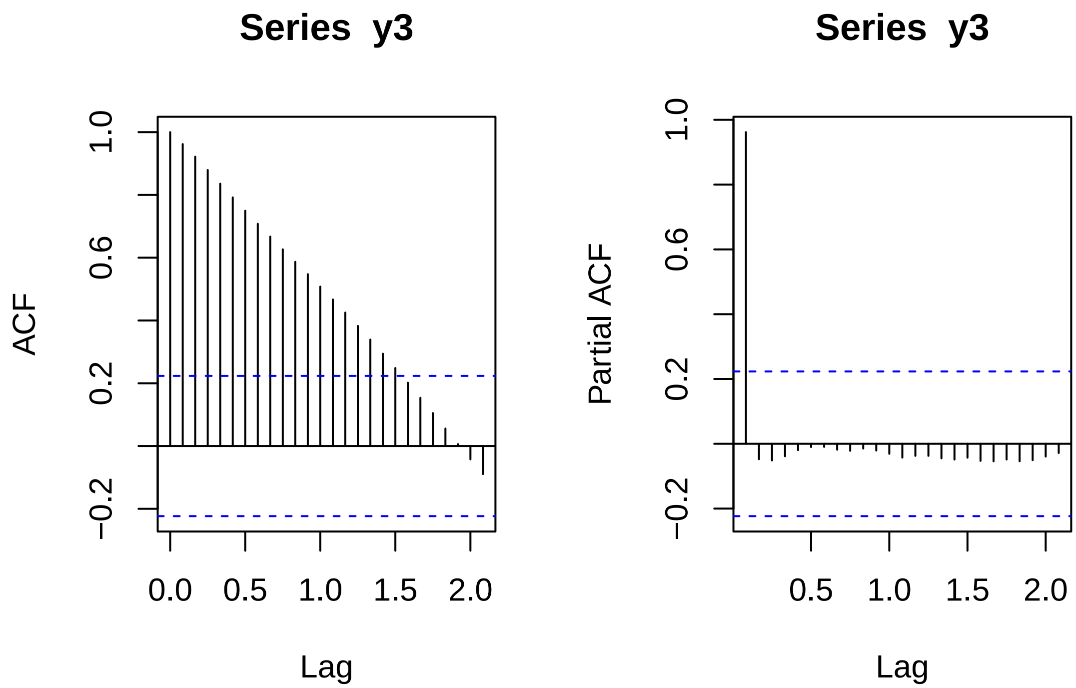
We see that for y_1 there is a clear seasonal trend and linear trend

```
par(mfcol = c(1, 2))
acf(y2, lag.max = 25)
pacf(y2, lag.max = 25)
```



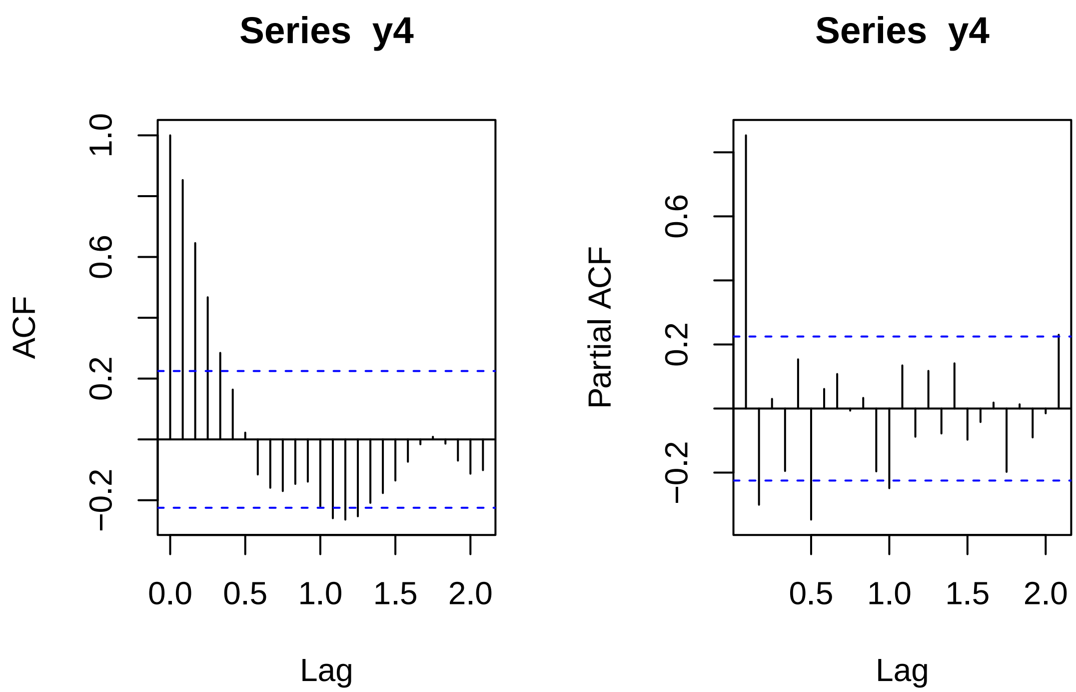
We see that for y_2 the seasonal and linear trend are removed

```
par(mfcol = c(1, 2))
acf(y3, na.action = na.pass, lag = 25)
pacf(y3, na.action = na.pass, lag = 25)
```



We see that for y_3 it has no seasonal trend but a linear trend.

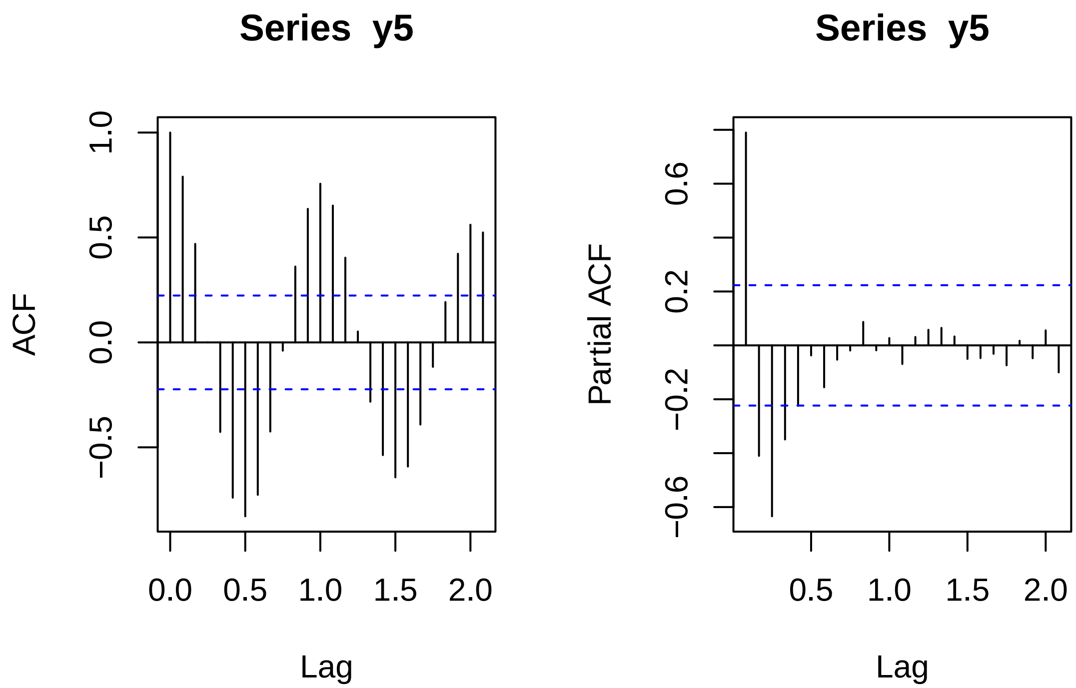
```
par(mfcol = c(1, 2))
acf(y4, na.action = na.pass, lag = 25)
pacf(y4, na.action = na.pass, lag = 25)
```



we see that though the first few lags (1, 2, 3, 4) are correlated, the lag decays quickly. There is no more seasonal and linear component.

```
par(mfcol = c(1, 2))
acf(y5, na.action = na.pass, lag = 25)
```

```
pacf(y5, na.action = na.pass, lag = 25)
```



We see that there is a linear and seasonal trend.