Lab 7

1. After inspecting the time series:

a) What is the time range (start and end) of this time series?

“month-year”
start: Jan-03
end: Dec-14

b) How many values are there in this time series?

144 value

c) What is the time interval separating between each two consecutive values in the time series? (monthly, weekly, daily, yearly, etc)

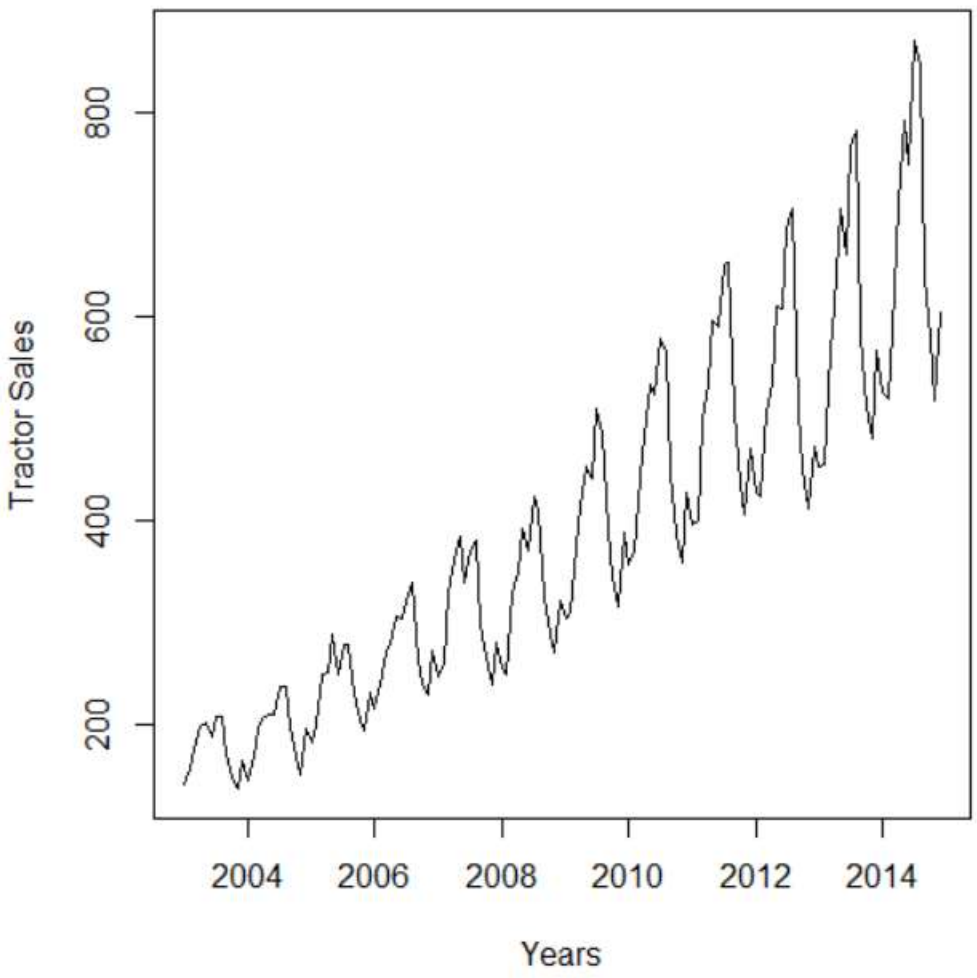
monthly

2. What does the parameter frequency mean ? Why did we set it to 12?

frequency: the number of observations per unit of time.
Because here our time unit is a year and the observations is monthly recorded.
Since one year contains 12 months so, frequency is set to 12.

3. **After visualizing the time series:**

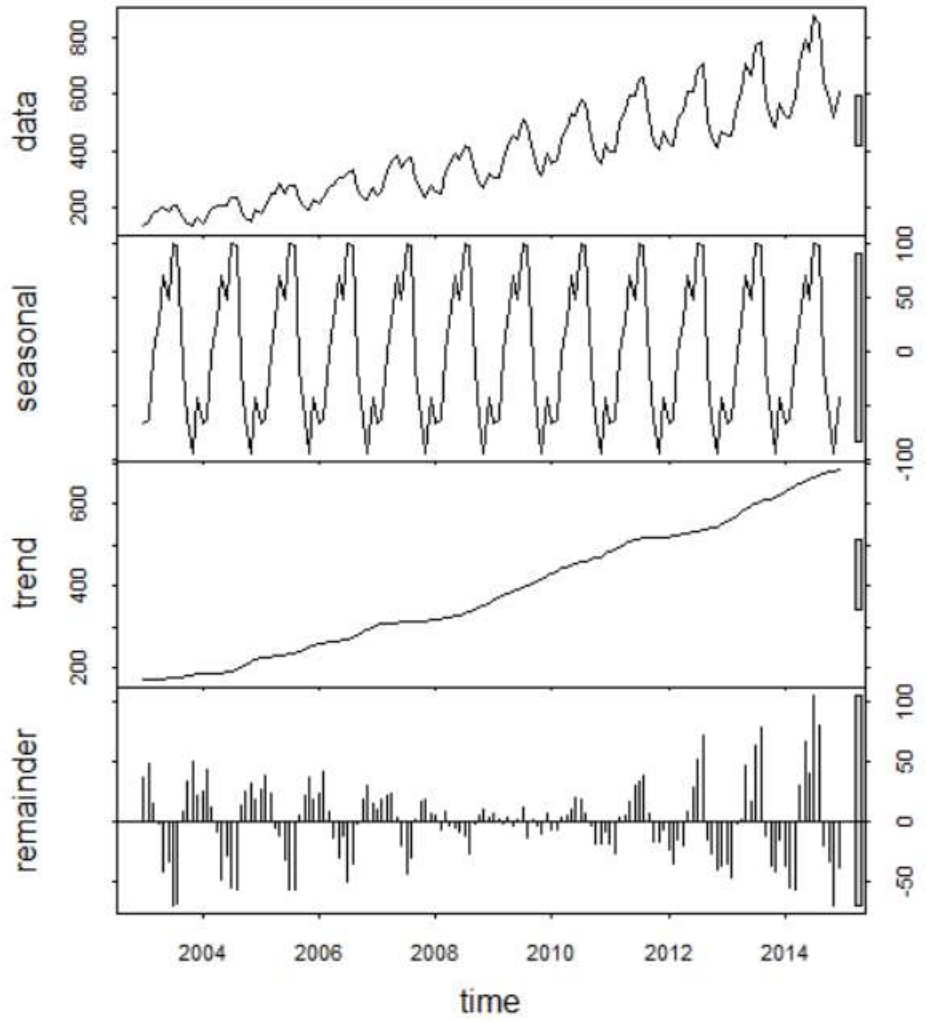
a) Add a neat plot of the generated time series



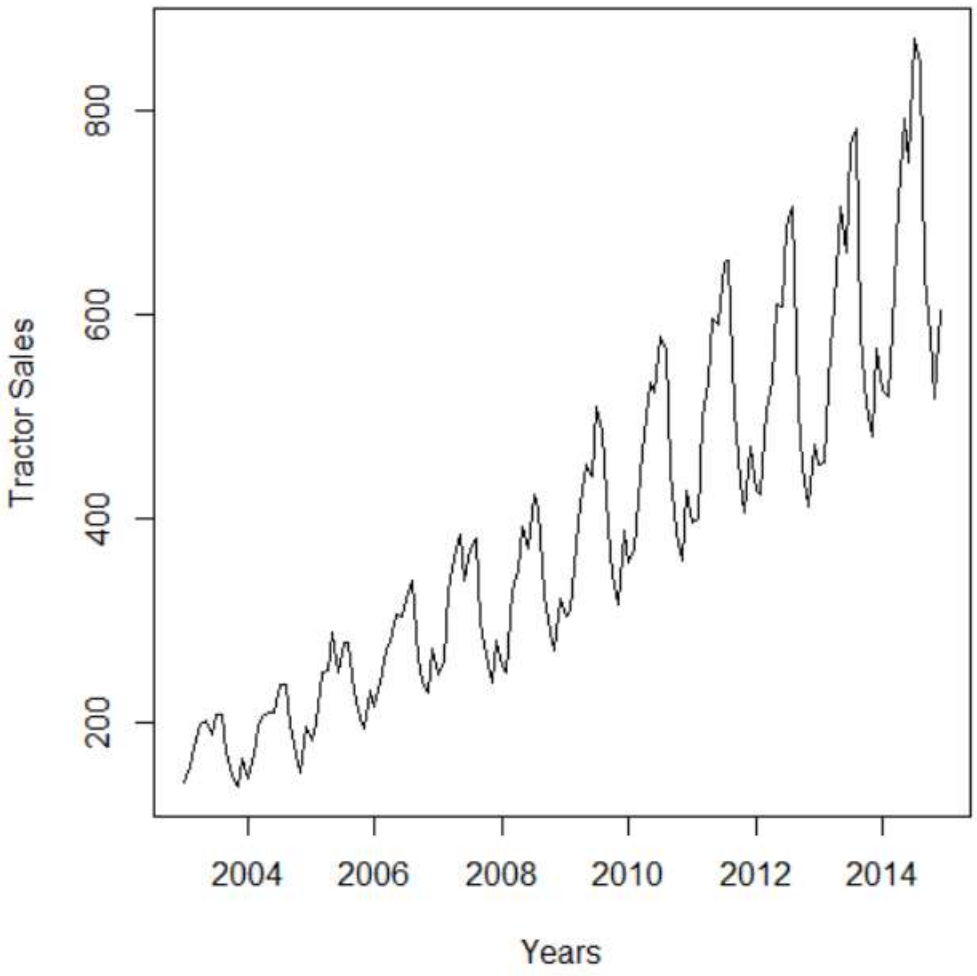
b) Do you think there is a trend in the time series? If yes, then what is the degree of the trend (i.e. is it linear, quadratic, .. etc)?
Yes, linear "observed from the graph"

c) Do you think there is a seasonality in the time series?
Yes

4. What does the function stl do? Add a neat plot of the plot generated
Decompose a time series into seasonal, trend and irregular component.



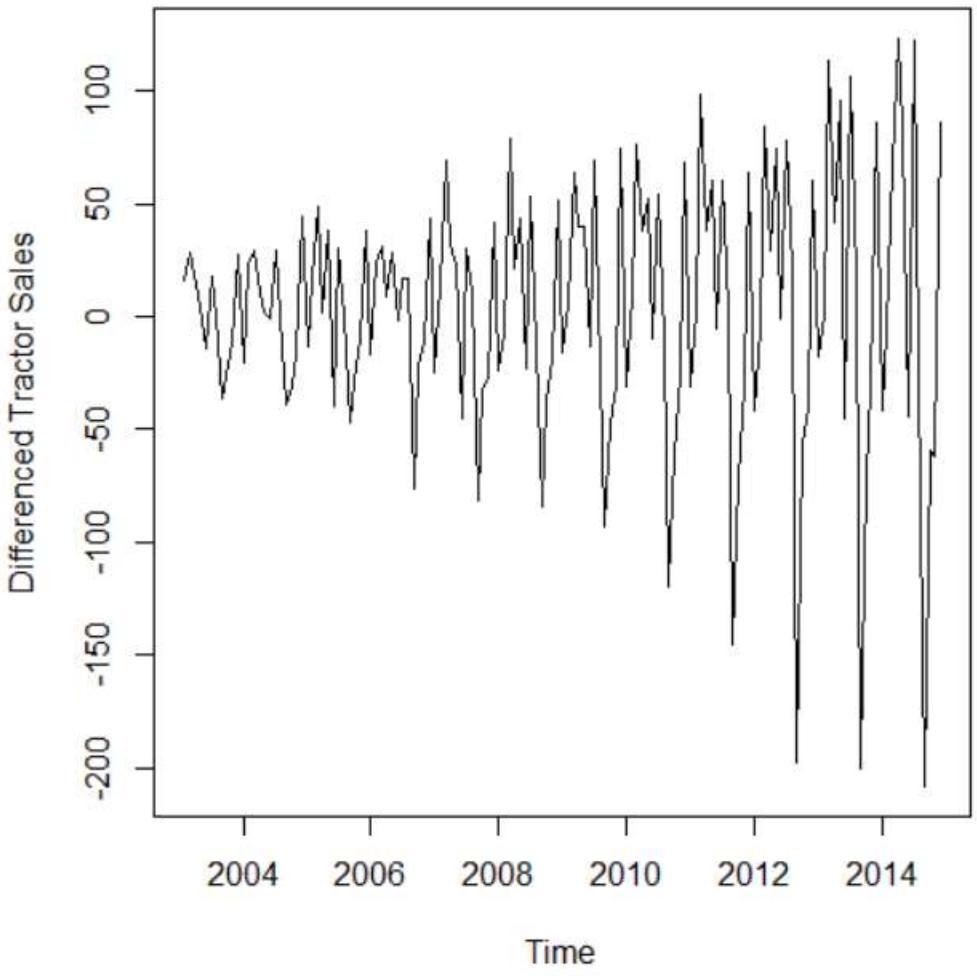
5. Back to the original time series:



a) What are the two conditions imposed on the mean and the variance of a time series to be stationary?
To be stationary, the mean and the variance shouldn't change over time

b) Is this time series stationary? Mention the reasons behind your answer.
No, there is trend and seasonality in this time series.
The mean increases over time.
The variance "the distance from mean" also increases over time.
There is also autocorrelation. "seasonality"

6. **After differencing the time series:**
a) Add a neat plot of the time series after differencing.



b) Does the time series become stationary? Are the two conditions of the mean and variance satisfied?

No, the time series doesn't become stationary

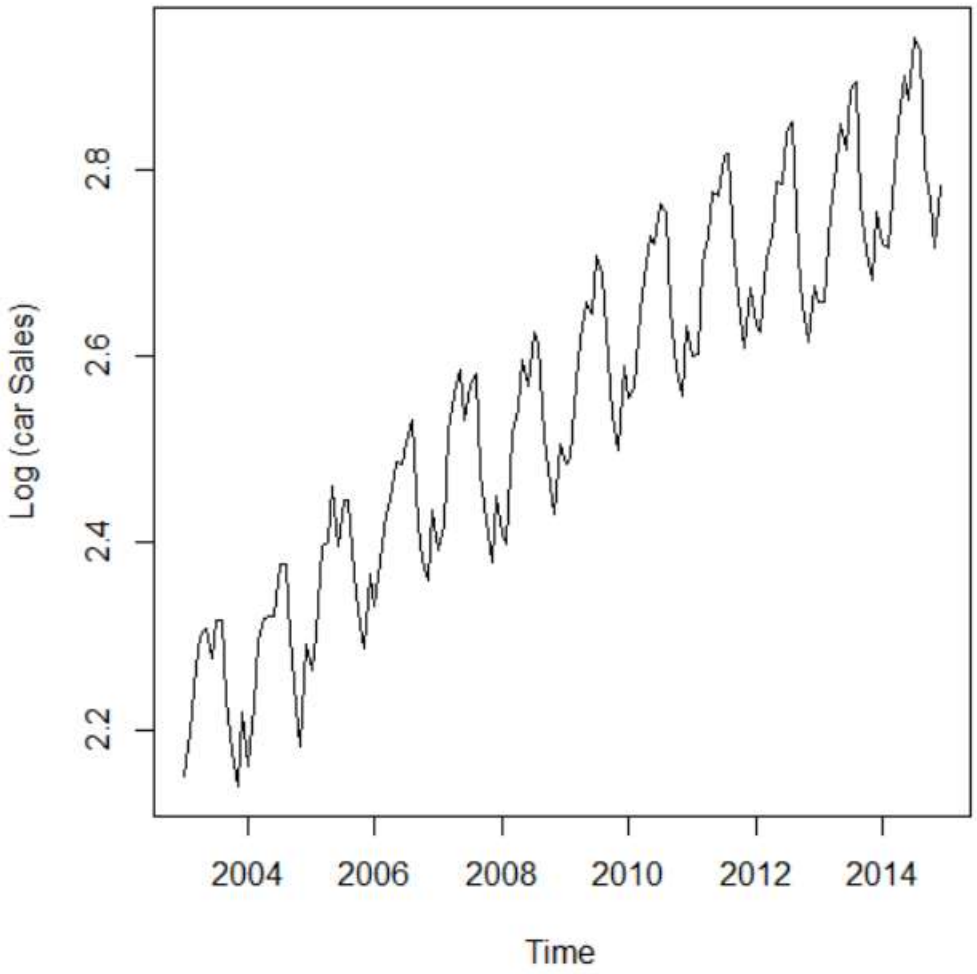
No, the variance still changes over time

c) If no, which of the two conditions is still not satisfied for a stationary time series?

The variance still changes over time

d) How does differencing help (not guarantee) to make a time series stationary?
It makes the mean constant.
It's like taking the derivative.
Since the time series was linear, taking difference once will make the mean become constant.

7. **After applying logarithm to the time series, compare visually (5) and (7) and you will know the answer to the following questions:**
a) Add a neat plot of the time series after applying logarithm.



b) Does the time series become stationary? Are the two conditions of the mean and variance satisfied?
No, it doesn't.
No, the mean still changes over time.

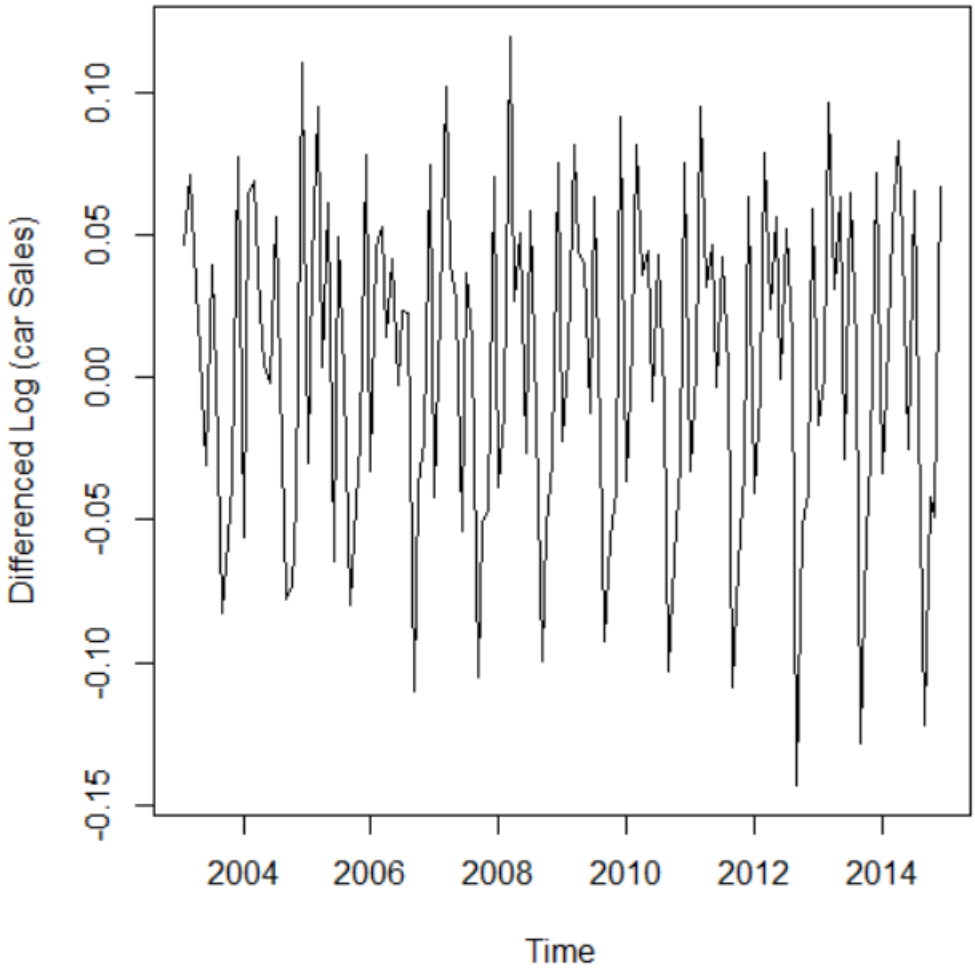
c) If no, which of the two conditions is still not satisfied for a stationary time series?

The mean still changes over time

d) How does applying logarithm help (not guarantee) to make a time series stationary? [This is a new piece of information never told in the lecture or the tutorial].
logarithms can help to stabilize the variance of a time series

8. **After applying both differencing and logarithm to the time series:**
a) Does the time series become stationary? Are the two conditions of the mean and variance satisfied?
Yes, the time series becomes stationary
Yes, the two conditions are satisfied

b) Add a neat plot of the final time series.



9. **After fitting an ARIMA Model with the logarithm of the time series:**
a) What are the two requirements of ARIMA (or ARMA) models on the time series data?
Requirement:

1- Detrended time series
2- Seasonally-adjusted time series

(Stationary sequence).
But for ARIMA we can let it make the detrending part by differencing.

b) Does the time series passed to the ARIMA model successfully hold the two requirements? Why?
No, There is still a trend in the time series "mean still changes over time"
and there is a seasonal effect in the time series so, we passed "seasnality = true" to the function-by default-.

c) Inspect the summary of the model.
ARIMA(0,1,1)(0,1,1)[12]
The output of the ARIMA model is: (p,d,q) (P,D,Q)[S]
What does the (p, d, q) mean?
p: number of autoregressive terms
d: number of differences
q: number of moving average terms

What do you think the (P, D, Q)[S] relate to? (no details are needed).
(P, D, Q) for the seasonal part of model.
[S] stands for number of periods in season. Here [12] is the number of months in year.

d) Do you think that ARIMA model achieved the two requirements of (9-a) internally? If yes, how did it happen briefly? You don't need to give any mathematical proofs or so. You just need to observe the ARIMA model output (9-c) and you will get it.
Yes, ARIMA did the differencing to detrend the time series by adding d=1 since the time series was linear, taking difference once will make the mean become constant
And it handled the seasonal effect through (P, D, Q)[S] by adding D = 1 as "seasonal = TRUE" is passed to the function.

e) What do you think will be more suitable for the case of forecasting the tractor sales, an autoregressive (AR) model or a moving average (MA) model? Why?
MA, the time series shows short term dependencies.

10. After changing trace = True:
a) How is the best model selected? What is the information criterion used in selecting the best model? [Mention only the name] [Check the help]
ARIMA(0,1,1)(0,1,1)[12] : -702.6033
AIC=-702.79 AICc=-702.6 BIC=-694.17

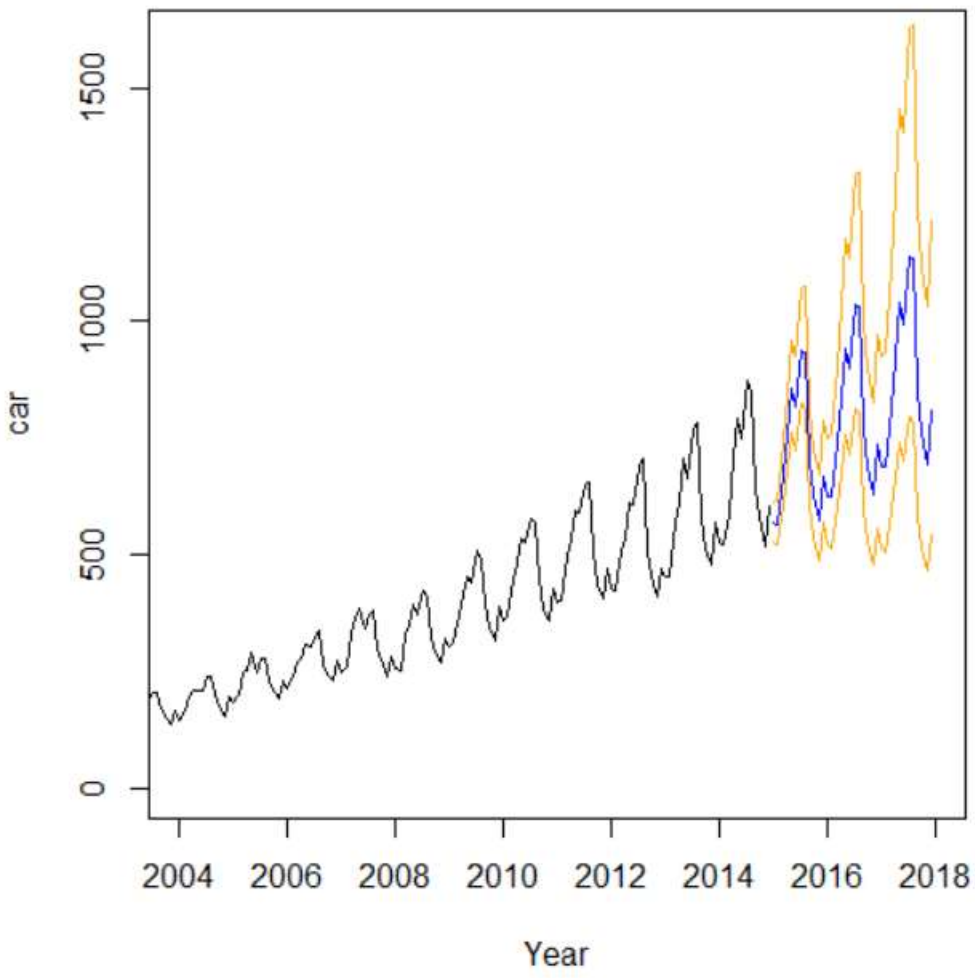
The model is selected using Information Criteria AICc

b) What other information criteria are there that can be used as well? [Mention only the names][Check the help]
AIC and BIC

c) Do we seek to get the minimum value or the maximum value of this criterion?
minimum

11. What is the meaning of n.head = 36?
The number of steps ahead for which prediction is required = 36 months = 3 upcoming years

12. After forecasting and plotting the future values
a) According to your observation, does this forecast work well?
Yes, it follows the same trend and season as previous years
b) Add a neat plot of the generated time series.

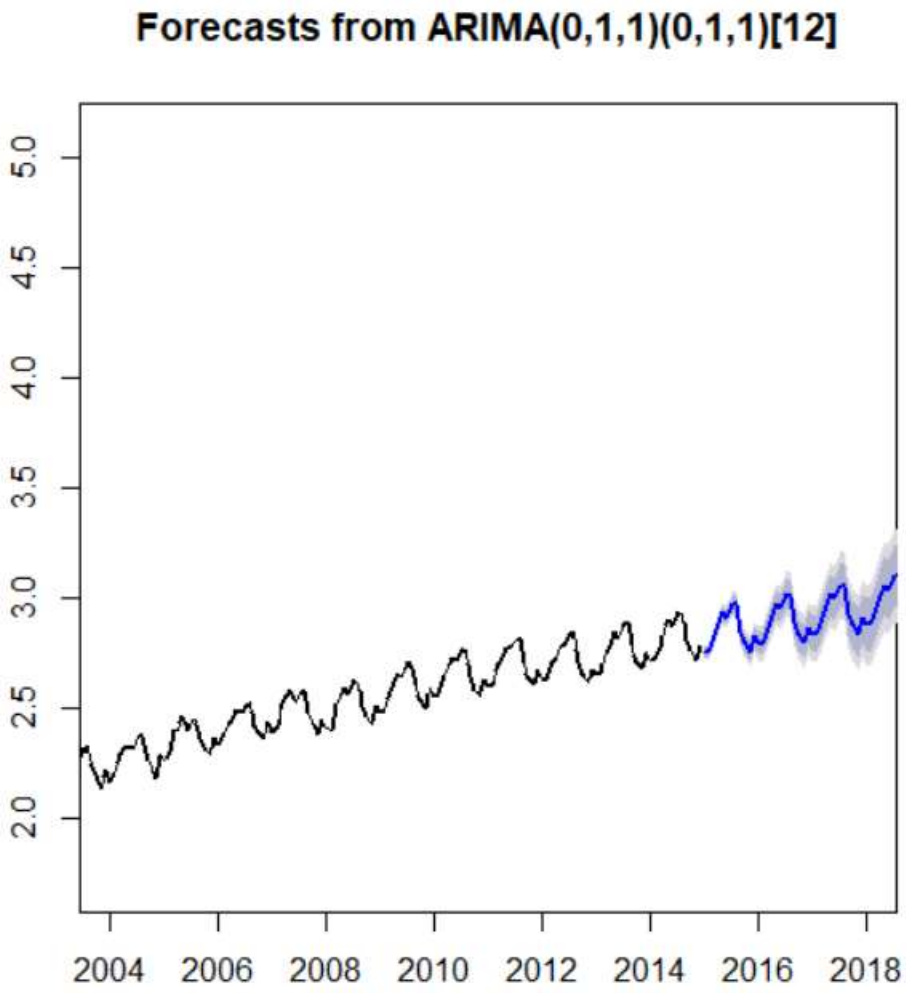


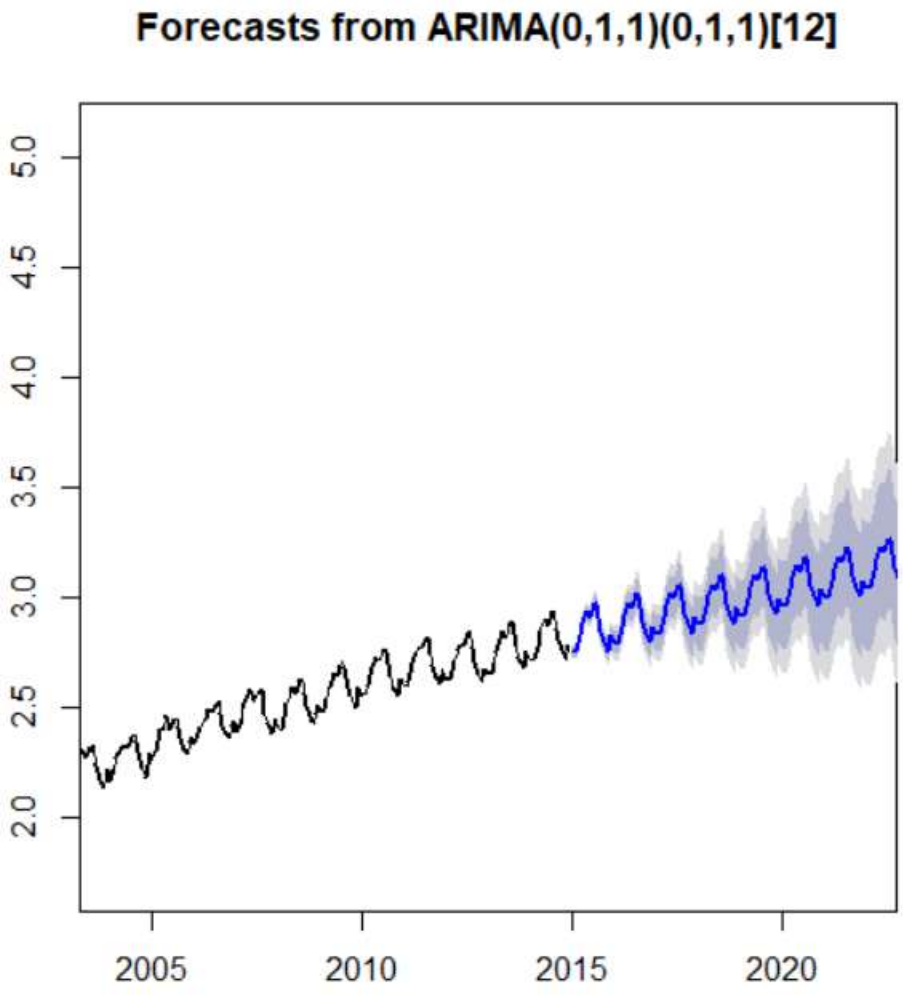
13. After using TSPred library:

a) Does this library generate a similar plot to (12)? Why?

No, because we took the log before passing the data to the ARIMA model
The generated graph is variance constant not like the actual data.

b) Add a neat plot of the generated time series.





14. What happened when we tried to forecast the tractor sales for an extended or longer time range? What do you notice?

The probability of error increases as we extend the time range.