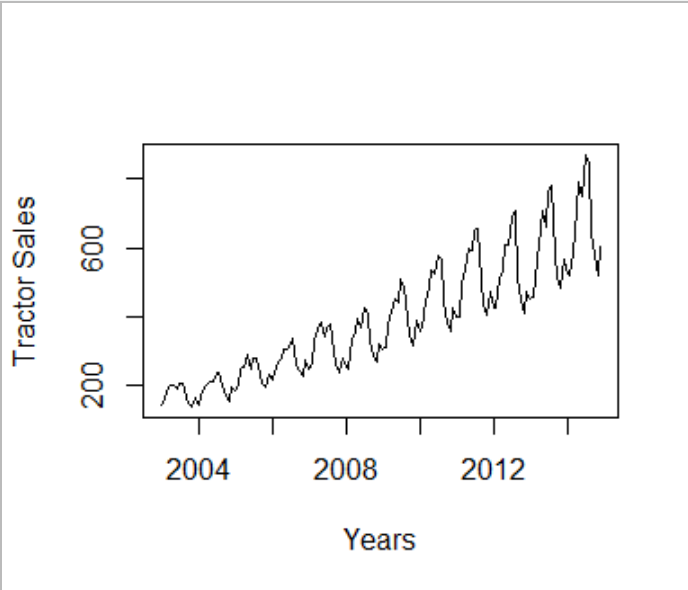
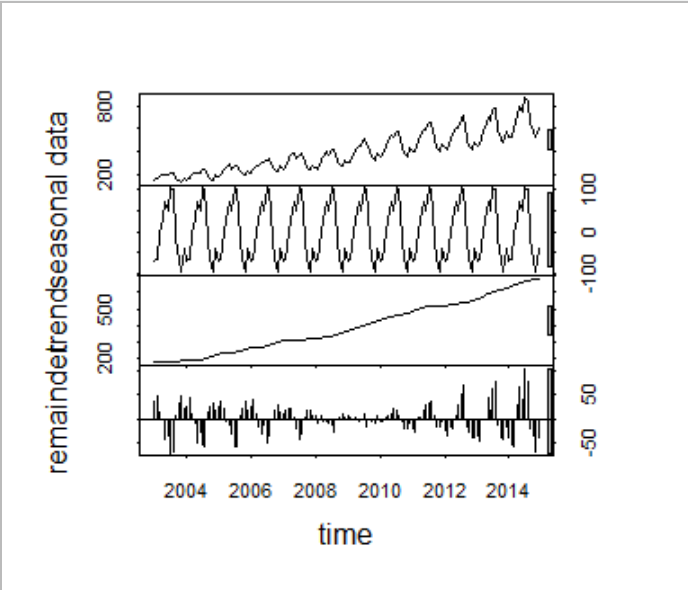
Time Series Requirement

* 1. Start: January 2003

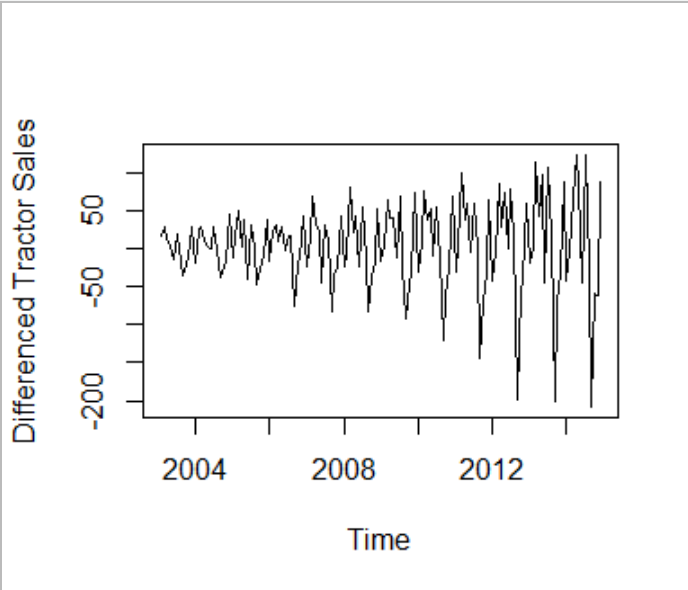
End: 2015

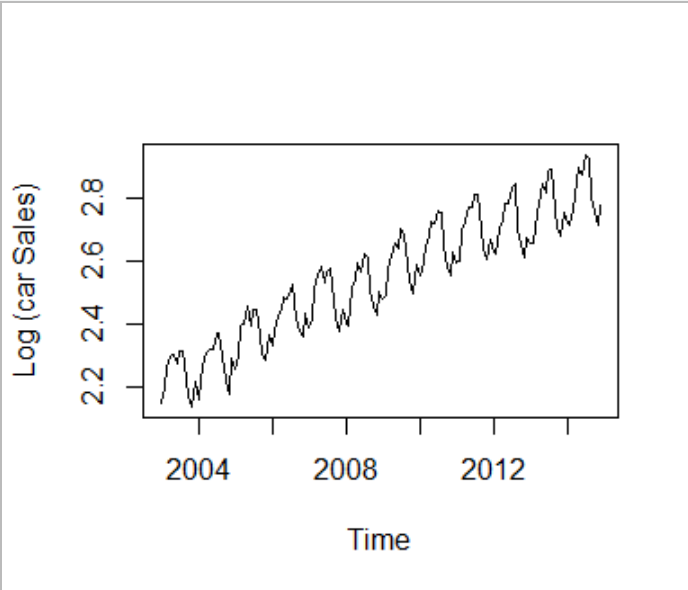
* 1. There are 144 values in the time series (144 data row)
  2. It is a monthly time interval, since the frequency is 12

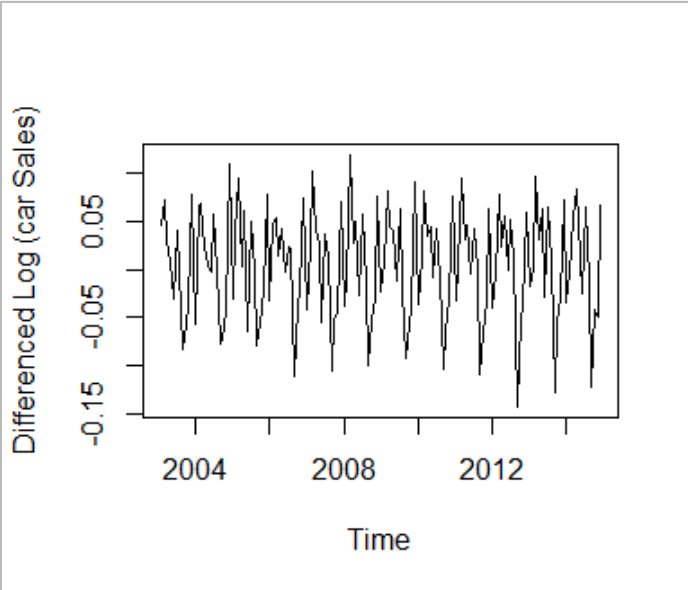
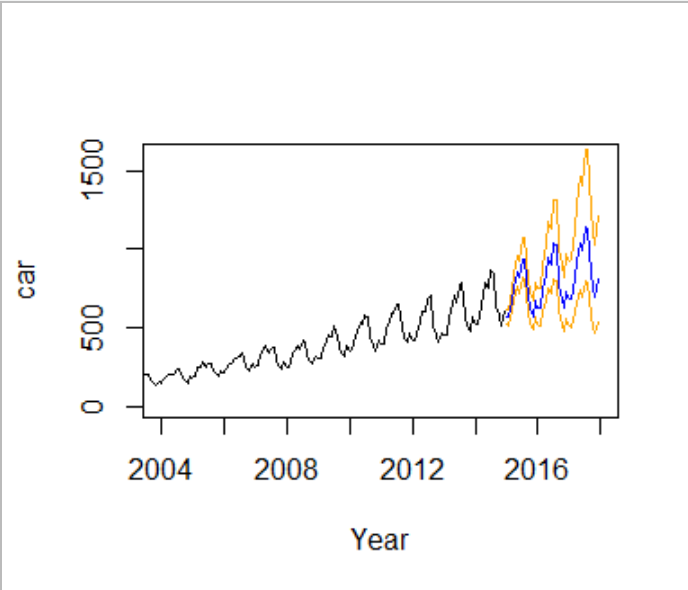
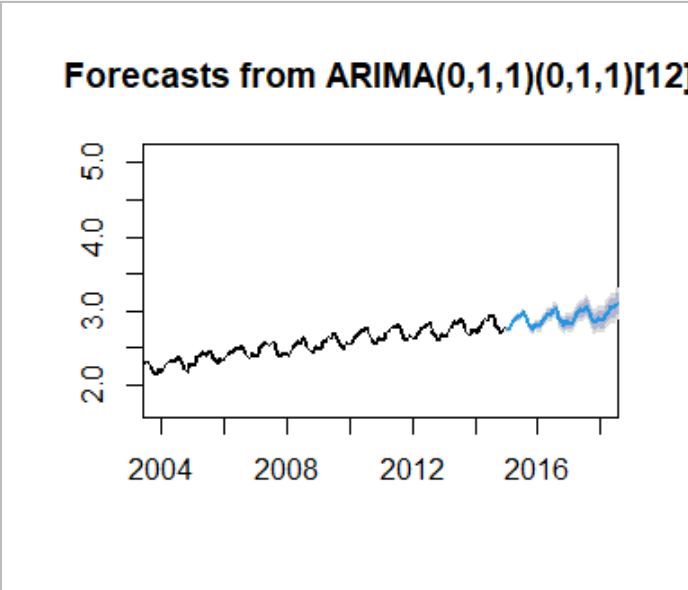
1. Frequency parameter means the number of observations per unit time. In order to represent data in years, a year contains 12 months, so setting the frequency value to 12 will show us data grouped per year.
   1. 
   2. Yes, there is a trend of linear degree.
   3. Yes, there is seasonality in the time series, a particular cycle repeats every year (every 12 months -> time period) which indicates seasonality in time series.
2. STL is a function which analyzes and decomposes the time series into trend, seasonality, and random fluctuations in the data (the irregular components).



* 1. No, the time series is not stationary,
  2. For the time series to be stationary, it should have a constant mean, variance and auto correlation. In this time series, the mean is changing over time and is shifted every year, so the variance is also shifted.



* 1. No, the time series did not become stationary.
  2. The data is de-trended, but no de-seasonalized, so the condition of constant variance is not satisfied yet.
  3. De-trending helps in reducing the data to have a constant mean. Since the data is linear, differencing it once is similar to differentiating it over time, this results in a constant mean, which is one of three requirements for the time series to be stationary.
  4. 
  5. No, the time series is still not stationary.
  6. The mean is still not constant after applying the log.
  7. Applying log to the time series converts the multiplicative (proportional-variance) seasonal pattern to an additive (constant-variance) seasonal pattern, which is one of the requirements of a time series to be stationary.

1. 
   1. Yes, the data is now stationary because both mean and variance are now constant, since the differencing helped in de-trending the data, and the log helped in de-seasonalizing the data.
   2. No, ARIMA requires stationary data as input (constant mean and variance), but the data entering the model here has only a constant variance since the log de-seasonalizes the data.
      1. In (p,d,q), **p** represents the last p-values the model is being fit to (for the Auto Regression part of the ARIMA model), **d** represents the number of differencing required to be done by the model (whether d=2 for quadratic trend or d=1 for linear trend and so on), **q** represents the irregularity in the last q-values the model is being fit to (for the Moving Average part of the ARIMA model).
      2. (P,Q,D) are the values the ARIMA model chooses for seasonal part of the model. [S] value here is the season frequency, which is monthly (12).
   3. Yes, if we observed the summary of the model, we would see that automatically the d value is set to 1, which means differencing by one level, since the data is linear. From this we can conclude that the model internally de-trended the data.
   4. I think moving average will be more suitable for tractor sales since they do not usually depend on the past sales for us to include auto regression. We can also notice that the model choose P=0 indicating that there are no auto regression terms.
   5. Best model selected is based on the AIC value
   6. Other criterions include AICc or BIC value
   7. The best model chosen was the one with the minimum negative value (largest in number but since it is negative so it is considered smaller). Models ranging from -653 to -702, the best model was the one with -702.
2. N.ahead = 36 means it should predict 36 observations (36/12 = 3 years).
   1. Yes, this is a reasonable forecast, since the increase in trend is also reasonable and the seasonality is also following the same pattern as the data observations (originally).
   2. 
   3. No, the plot is different since the predictions were logged, unlike the one in part (12).
   4. 
3. Drawing the forecast for a longer period of time resulted in poor predictions (indicated by a larger shaded area above and below the graph). This is because forecasting when done for long intervals will be less accurate since it has less values to forecast on, and it predicts on the already predicted values (which may have some errors), leading to even greater error as shown on the graph.