**Visualizing Data:**

*coaldata <- c (47730,46704,41535,41319,36962,32558,31995,32993,44834,29883,39611,*

*40099,38051,36927,37272,39457,38097,40226,43589,39088,39409,37226,*

*34421,34975,32710,31885,32106,30029,29501,31620,34205,32153,32764,*

*33230,35636,35550,34529,37498,37229,36021,38281,36676,44541,40850,*

*38404,37575,41476,42267,43062,45036,43769,42298,44412,40498,37830,*

*42294,38330,43554,42579,36911,42541,42430,43465,44468,43597,40774,*

*42573,41635,39030,41572,37027,34732,36817,34295,33218,32034,31417,*

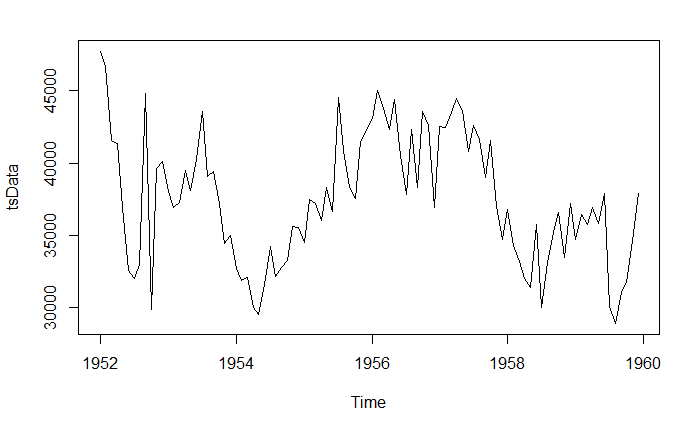
*35719,30001,33096,35196,36550,33463,37195,34748,36461,35754,36943,*

*35854,37912,30095,28931,31020,31746,34613,37901)*

*coaldata*

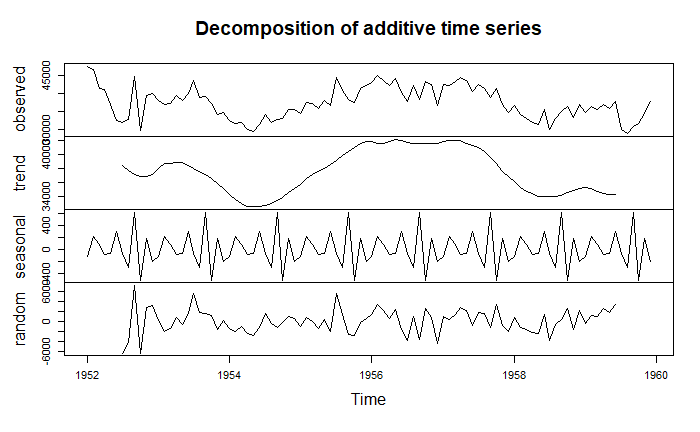
*tsData = ts(coaldata,start = c(1952,1), frequency = 12)*

*plot(tsData)*



*components.ts = decompose(tsData)*

*plot(components.ts)*



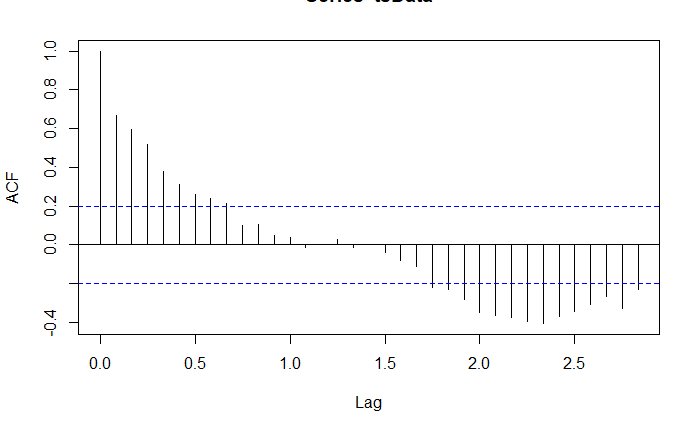
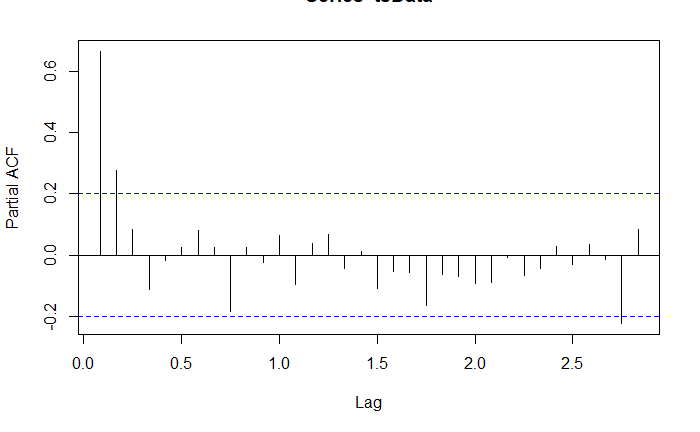
Here we get 4 components:

* Observed – the actual data plot
* Trend – the overall upward or downward movement of the data points
* Seasonal – any monthly/yearly pattern of the data points
* Random – unexplainable part of the data

Checking the ACF and PACF plots of the data to determine the order of the model to be used.

*acf(tsData,lag.max=34)*

*pacf(tsData, lag.max=34)*

**Fitting a model:**

We see here that the ACF is exponentially decreasing with lag, and the PACF is significant till 2 values of lag. In such a case we use the AR(2) model.

Although comparing AR1 and AR2 AIC and BIC values to confirm:

*fitARIMA <- arima(tsData, order=c(2,0,0),method="ML")*

*fitARIMAR2 <- arima(tsData, order=c(1,0,0),method="ML")*

*AIC(fitARIMAR1)*

*[1] 1822.801*

*BIC(fitARIMAR1)*

*[1] 1833.059*

*AIC(fitARIMAR2)*

*[1] 1830.813*

*BIC(fitARIMAR2)*

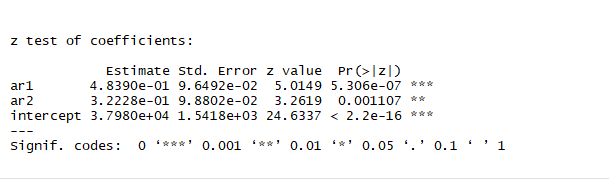
*[1] 1838.506*

Here again we can confirm that the AR(2) model is a better fit.

*install.packages('FitAR')*

*library(lmtest)*

*coeftest(fitARIMA)*

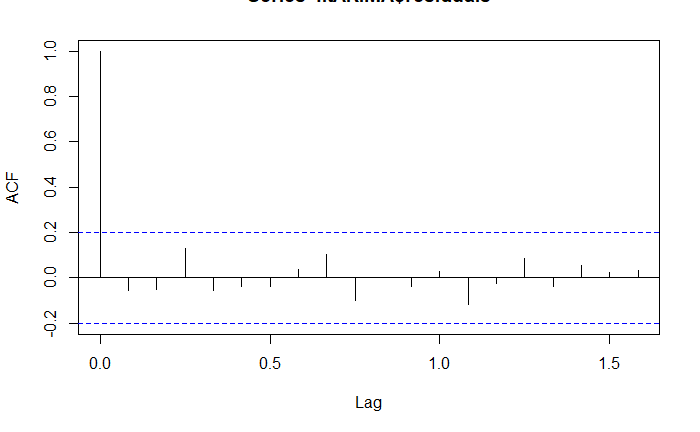
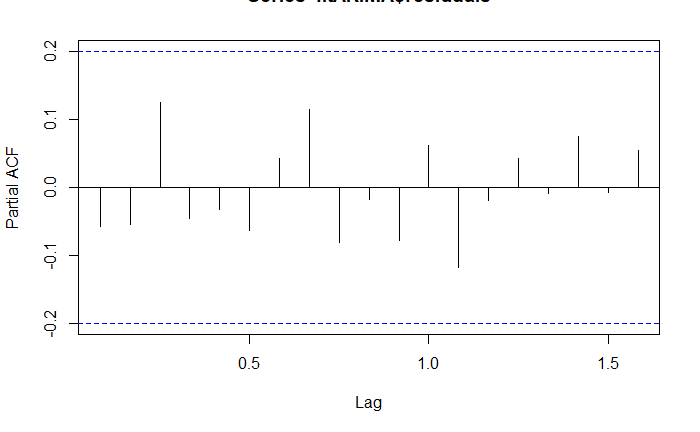


**Residual Diagnostics:**

We need to check if the residual corresponds to the white noise.

*acf(fitARIMA$residuals)*

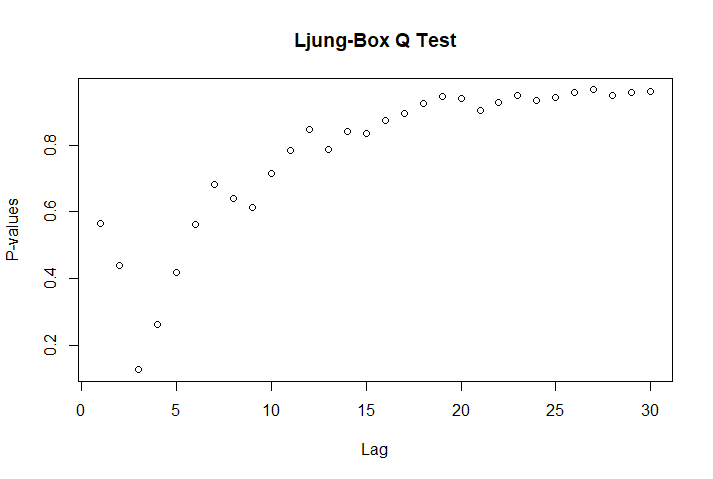
*pacf(fitARIMA$residuals)*

The ACF of the residuals shows no significant autocorrelations.

*boxresult<-LjungBoxTest (fitARIMA$residuals,k=2,StartLag=1)*

*plot(boxresult[,3],main= "Ljung-Box Q Test", ylab= "P-values", xlab= "Lag")*

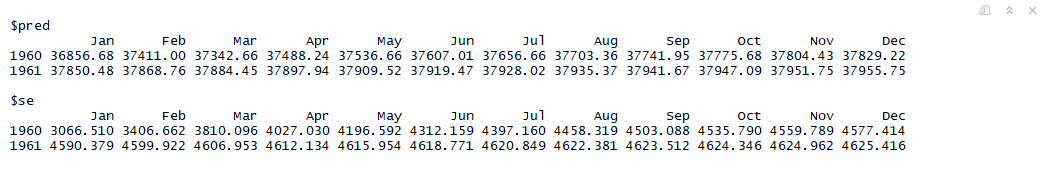


The p-values for the Ljung-Box Q test all are well above 0.05, indicating “non-significance.”

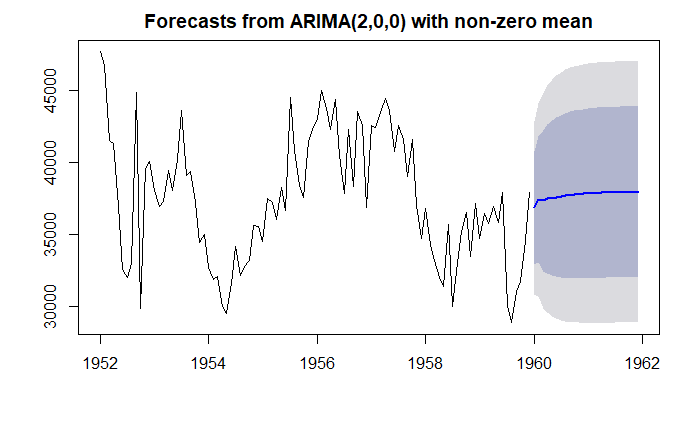
As all the graphs are in support of the assumption that there is no pattern in the residuals, we can go ahead and calculate the forecast.

**Predicting Future values:**

*predict(fitARIMA,n.ahead = 24)*



*plot(forecast(fitARIMA,h = 24))*



The forecasts are shown as a blue line, with the 80% prediction intervals as a dark shaded area, and the 95% prediction intervals as a light shaded area.