Name – Manish Sanjay Talekar

QMB Assignment 6

U54500765

#preprocessing

**> master\_data <- read\_xlsx("6304 Time Series Assignment Data.xlsx")**

**> colnames(master\_data)=tolower(make.names(colnames(master\_data)))**

**> master\_data$item=seq(1:nrow(master\_data))**

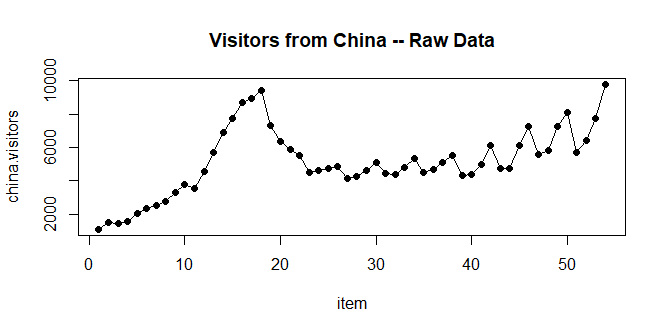
**> attach(master\_data)**

#Analysis

#q1

**> plot(item,china.visitors,type="o",pch=19,**

**+ main="Visitors from China -- Raw Data")**



#q2

**> visit\_out <- lm(china.visitors~item, data = master\_data)**

**> summary(visit\_out)**

**Call:**

**lm(formula = china.visitors ~ item, data = master\_data)**

**Residuals:**

**Min 1Q Median 3Q Max**

**-2273.0 -1071.4 -436.2 739.3 4929.8**

**Coefficients:**

**Estimate Std. Error t value Pr(>|t|)**

**(Intercept) 3294.93 477.58 6.899 7.17e-09 \*\*\***

**item 67.07 15.11 4.439 4.73e-05 \*\*\***

**---**

**Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1**

**Residual standard error: 1730 on 52 degrees of freedom**

**Multiple R-squared: 0.2748, Adjusted R-squared: 0.2609**

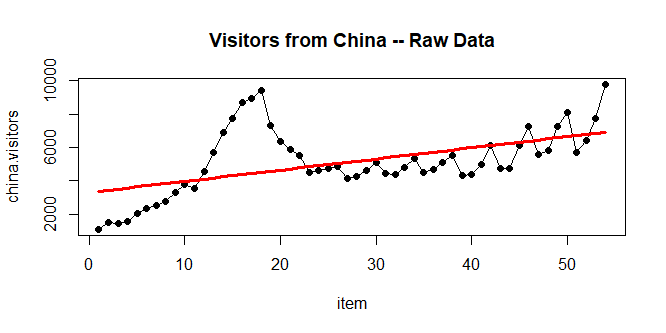
**F-statistic: 19.71 on 1 and 52 DF, p-value: 4.728e-05**

#q3

**> plot(item,china.visitors,type="o",pch=19,**

**+ main="Visitors from China -- Raw Data")**

**> points(visit\_out$fitted.values,type="l",lwd=3,col="red")**



#q4

**> durbin.out=durbinWatsonTest(visit\_out)**

**> durbin.out**

**lag Autocorrelation D-W Statistic p-value**

**1 0.8219683 0.2701599 0**

**Alternative hypothesis: rho != 0**

INTERPRETATION –

* Based on the above p value we can conclude that the null hypothesis can be rejected. Which means that there is autocorrelation between the error points.
* Since the D-W Statistic value is close to zero we can conclude that there is positive serial autocorrelation.

#q5

#Making Seasonal Indices

**> indices=data.frame(quarter=1:4,average=0,index=0)**

**> for(i in 1:4) {**

**+ count=0**

**+ for(j in 1:nrow(master\_data)) {**

**+ if(i==master\_data$quarter[j]) {**

**+ indices$average[i]=indices$average[i]+master\_data$china.visitors[j]**

**+ count=count+1**

**+ }**

**+ }**

**+ indices$average[i]=indices$average[i]/count**

**+ indices$index[i]=indices$average[i]/mean(master\_data$china.visitors)}**

#Deseasonalizing the original data

**> for(i in 1:4){**

**+ for(j in 1:nrow(master\_data)){**

**+ if(i==master\_data$quarter[j]){**

**+ master\_data$deseason.visitors[j]=**

**+ master\_data$china.visitors[j]/indices$index[i] }}}**

#q6

**> devisit\_out <- lm(deseason.visitors~item, data=master\_data)**

**> summary(devisit\_out)**

**Call:**

**lm(formula = deseason.visitors ~ item, data = master\_data)**

**Residuals:**

**Min 1Q Median 3Q Max**

**-2339.6 -892.8 -595.9 457.4 4830.8**

**Coefficients:**

**Estimate Std. Error t value Pr(>|t|)**

**(Intercept) 3358.33 459.19 7.314 1.57e-09 \*\*\***

**item 64.77 14.53 4.458 4.43e-05 \*\*\***

**---**

**Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1**

**Residual standard error: 1664 on 52 degrees of freedom**

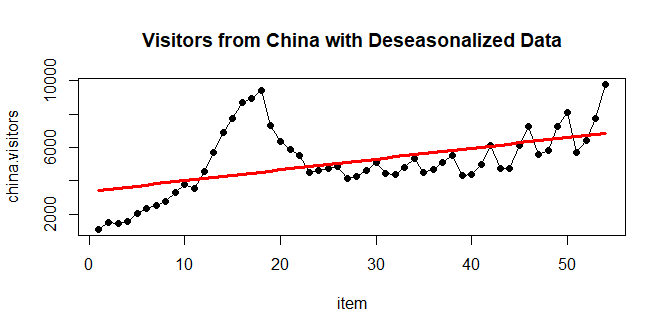
**Multiple R-squared: 0.2765, Adjusted R-squared: 0.2626**

**F-statistic: 19.88 on 1 and 52 DF, p-value: 4.434e-05**

**> plot(item,china.visitors,type="o",pch=19,**

**+ main="Visitors from China with Deseasonalized Data")**

**> points(devisit\_out$fitted.values,type="l",lwd=3,col="red")**



#q7

#Reseasonalizing Forecasts

**> master\_data$deseason.forecast=devisit\_out$fitted.values**

**> for(i in 1:4){**

**+ for(j in 1:nrow(master\_data)){**

**+ if(i==master\_data$quarter[j]){**

**+ master\_data$reseason.forecast[j]=master\_data$deseason.forecast[j]\***

**+ indices$index[i]**

**+ }**

**+ }**

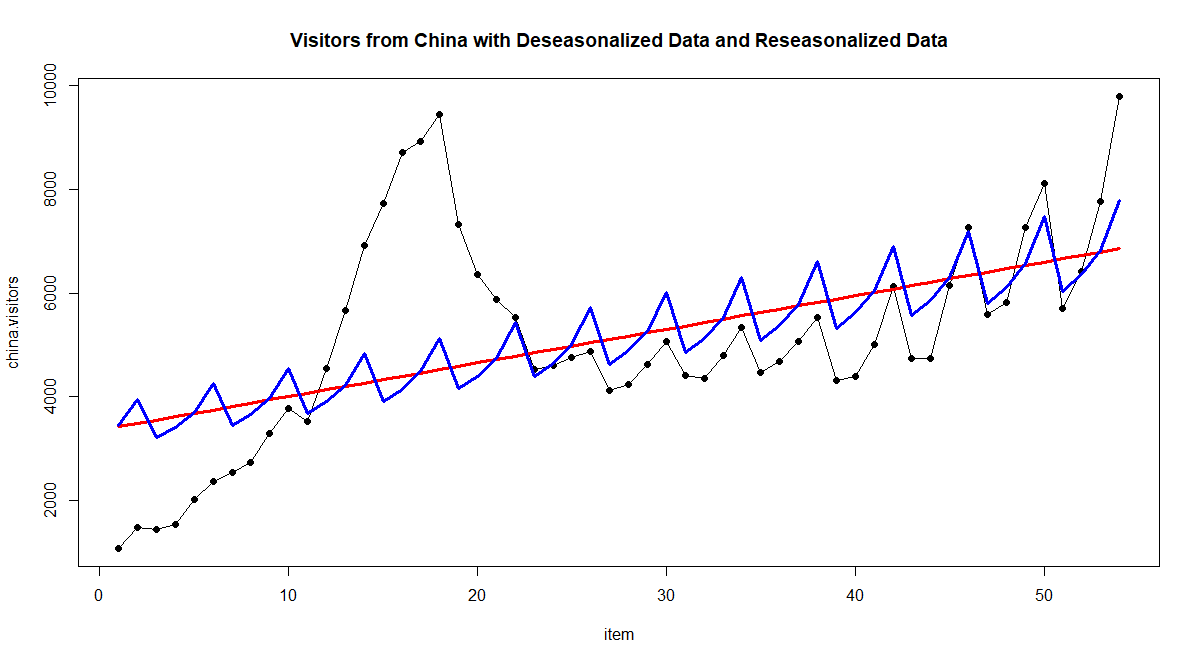
**+ }**

**> plot(item,china.visitors,type="o",pch=19,**

**+ main="Visitors from China with Deseasonalized Data and Reseasonalized Data")**

**> points(devisit\_out$fitted.values,type="l",lwd=3,col="red")**

**> points(master\_data$reseason.forecast,type="l",lwd=3,col="blue")**



INTERPRETATION – Based on the above plot, residual standard error, r-squared and adjusted r-squared values, I can conclude that neither the deseasonalized nor the reseasonalized data appear to have a better fit to the original data.

#q8

**> master\_data$error=master\_data$china.visitors-master\_data$reseason.forecast**

**> master\_data$stdzd.error=scale(master\_data$error)**

#Plots

**> par(mfrow=c(2,2))**

**> plot(master\_data$china.visitors,master\_data$error,pch=19,type="o",**

**+ xlab="Time Period",ylab="Error",**

**+ main="Reseasonalized Forecasts -- Errors",**

**+ sub="By Sequence")**

**> abline(0,0,col="red",lwd=3)**

**> plot(master\_data$china.visitors,master\_data$stdzd.error,type="o",pch=19,**

**+ main="Reseasonalized Forecasts -- Standardized Errors",**

**+ xlab="Time Period",ylab="Standardized Errors",**

**+ sub="By Sequence")**

**> abline(0,0,col="red",lwd=3)**

**> #Plot by china.visitors**

**> plot(master\_data$china.visitors,master\_data$error,pch=19,**

**+ main="Reseasonalized Forecasts -- Errors",**

**+ xlab="china.visitors",ylab=" Errors",**

**+ sub="By china.visitors")**

**> abline(0,0,col="red",lwd=3)**

**> plot(master\_data$china.visitors,master\_data$stdzd.error,pch=19,**

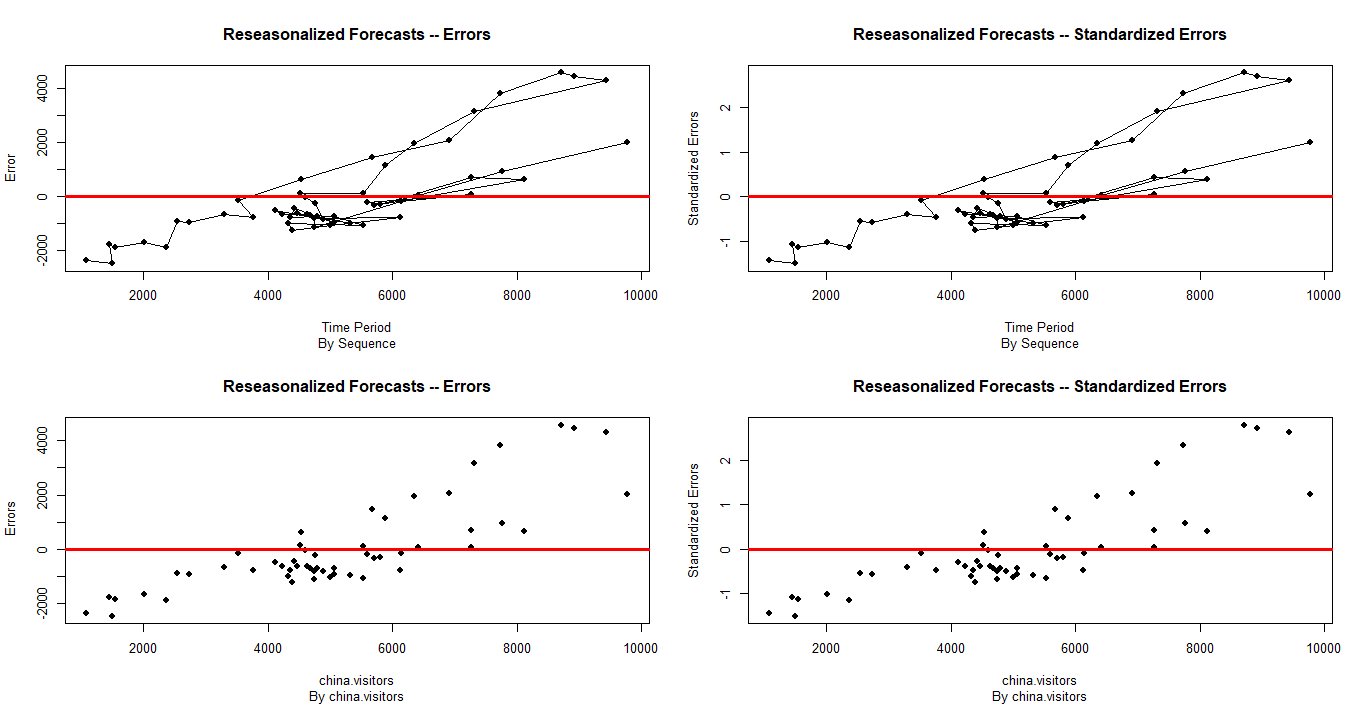
**+ main="Reseasonalized Forecasts -- Standardized Errors",**

**+ xlab="china.visitors",ylab="Standardized Errors",**

**+ sub="By china.visitors")**

**> abline(0,0,col="red",lwd=3)**

**> par(mfrow=c(1,1))**



INTERPRETATION – Based on above graphs there seems to be a pattern among the standardized errors. They seem to appear somewhat linear in nature.