MATH1318 Semester 1, 2020

Code

Assignment 1

Ancy Rex(S3760058)

Setup

Install and load the necessary packages :

Hide

Hide

```
library(TSA)
library(tseries)
library(forecast)
```

Read ozone layer thickness Data

Read the Ozone layer thickness data using an appropriate function.

Hide

Hide

```
# Reading the Ozone layer thickness data.
ozone <- read.csv("/Users/ancy_rex/Documents/S3\ /Time\ Series\ Analysis/Assignmen
ts/Assignment\ 1/data1.csv", header=FALSE)
#class of dataset
class(ozone)
```

```
[1] "data.frame"
```

CHAPTER 1

Introduction

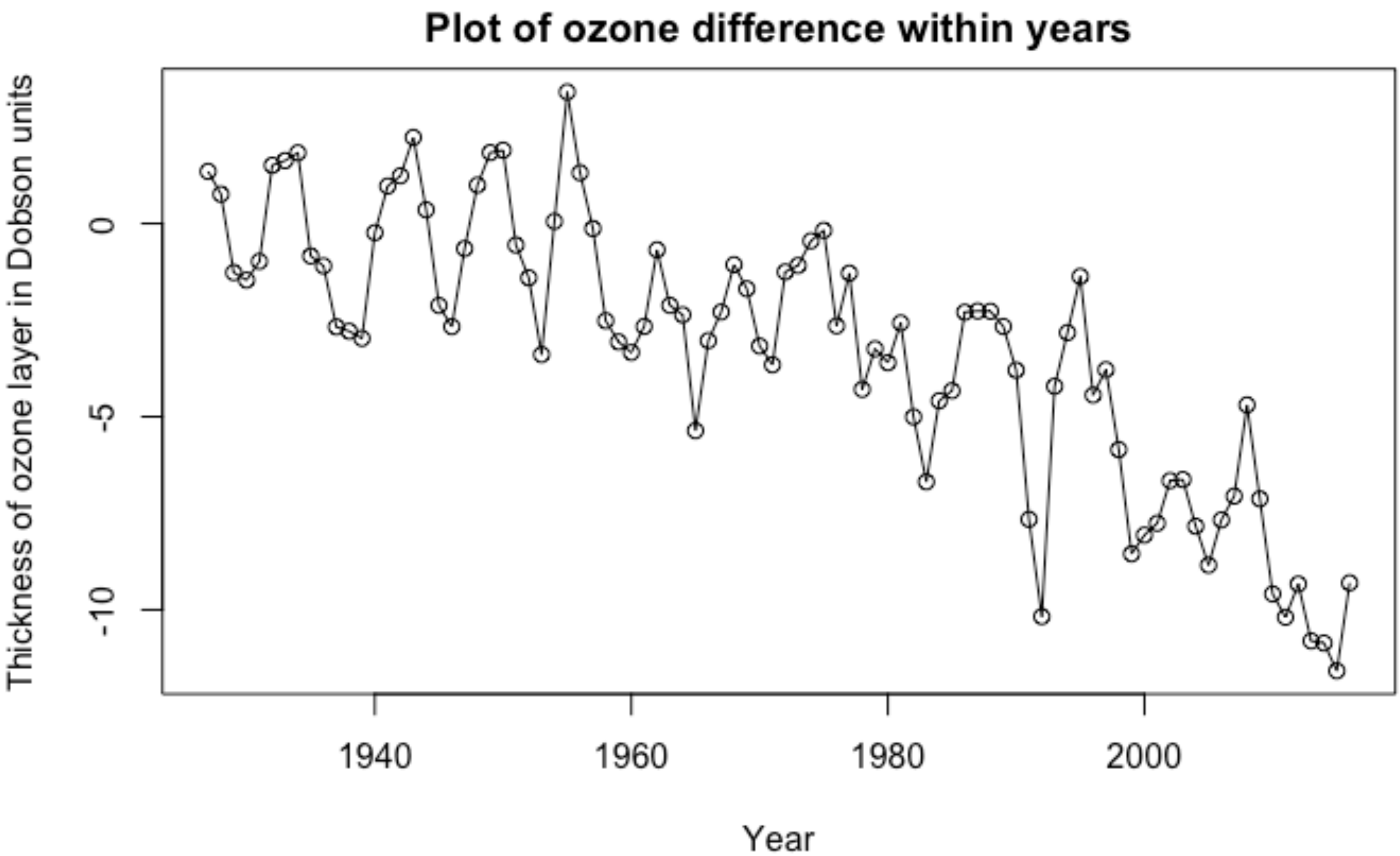
The dataset contains the changes in the thickness of ozone layer within 1927 to 2016 in Dobson units. The values in the dataset represents the thickess and if they are positive there is a increase in thickness and it decreases when the values are negative.

Plotting the ozone layer thickness in the years

Hide

Hide

```
rownames(ozone) <- seq(from=1927 , to=2016)
ozone <- ts(as.vector(ozone), start=1927 , end=2016)
#plotting time series graph
plot(ozone ,type='o',xlab='Year',ylab='Thickness of ozone layer in Dobson units',m
ain = "Plot of ozone difference within years")
```



The plot does not show change in the variance within the years but depicts steady decreasing fluctuations in the trend with no seasonability.

Linear Trend model

Estimating the slope and intercept by least-squares regression by considering this as a linear time trend

Hide

Hide

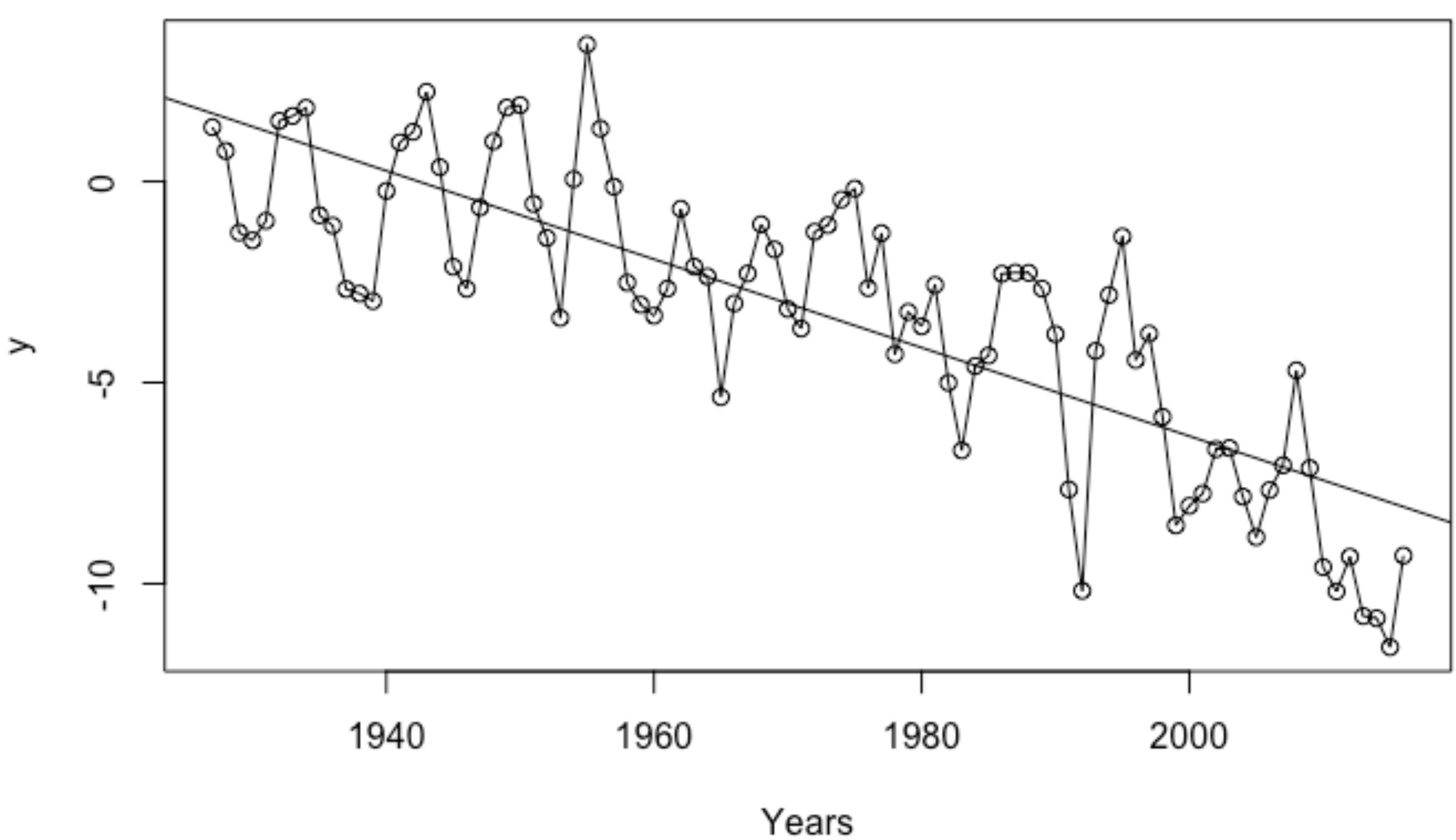
```
#Linear trend model
model.ozone.ln = lm(ozone~time(ozone))
summary(model.ozone.ln)
```

```
Call:
lm(formula = ozone ~ time(ozone))

Residuals:
    Min      1Q  Median      3Q     Max 
-4.7165 -1.6687  0.0275  1.4726  4.7940 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) 213.720155  16.257158   13.15   <2e-16 ***
time(ozone)  -0.110029   0.008245  -13.34   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.032 on 88 degrees of freedom
Multiple R-squared:  0.6693,	Adjusted R-squared:  0.6655 
F-statistic: 178.1 on 1 and 88 DF,  p-value: < 2.2e-16
```

Hide

Hide

```
plot(ozone,type='o',ylab='y',xlab='Years')
abline(model.ozone.ln)
```



The slope and intercepts of the above graph are 213.72 and -0.11 respectively. The model has adjusted R-square value of 0.6655 and so giving us 66% of the variance and significantly higher variables and intercept coefficients. The plot also gives us a slope that is statistically significant at 5% significance level.

Quadractic Trend model

Estimating the slope and intercept by least-squares regression by quadractic time trend

Hide

Hide

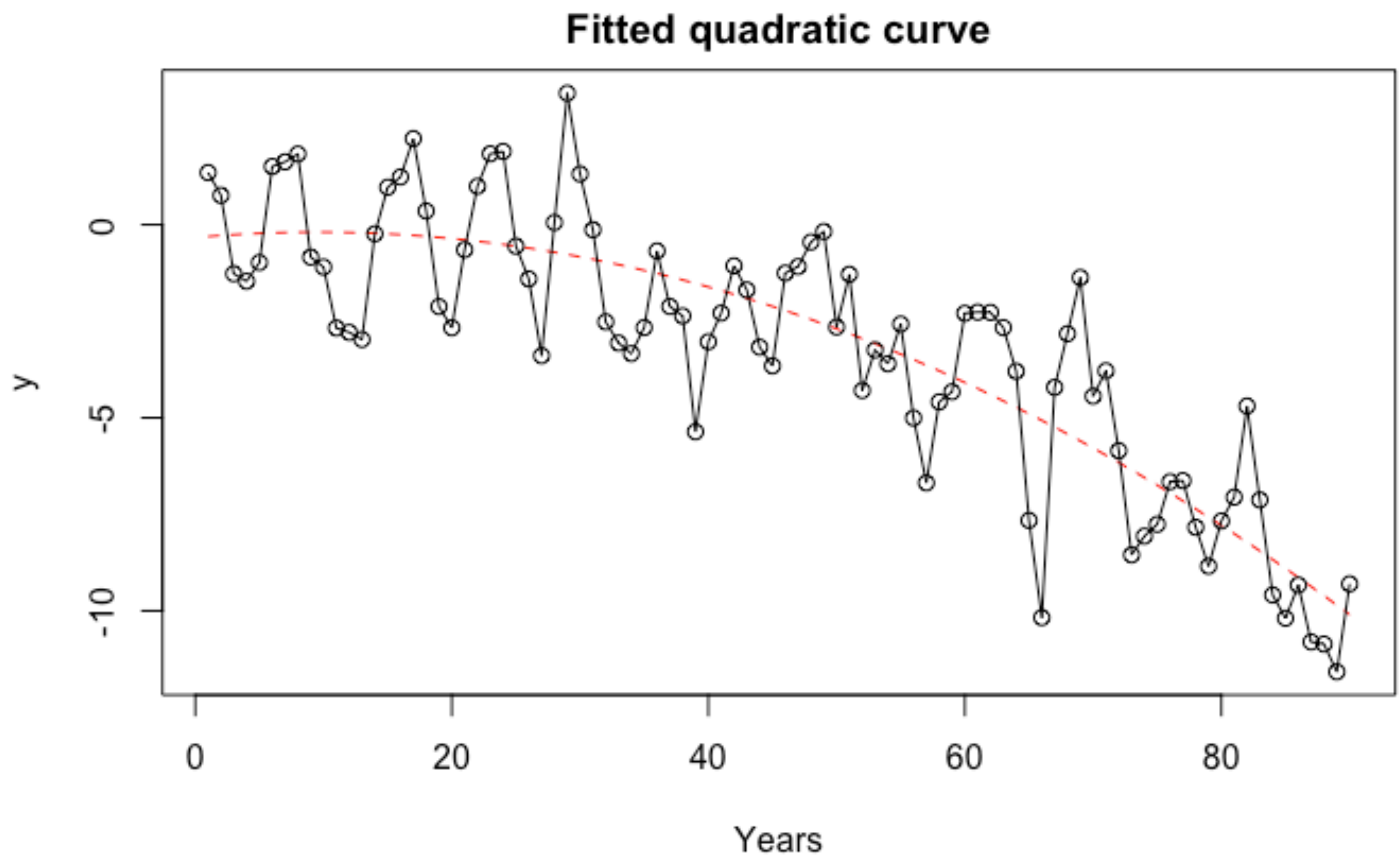
```
#quadractic time trend
t = time(ozone)
t2 = t^2
model.ozone.qa = lm(ozone~ t + t2)
summary(model.ozone.qa)
```

```

Call:
lm(formula = ozone ~ t + t2)

Residuals:
    Min      1Q  Median      3Q     Max 
-5.1062 -1.2846 -0.0055  1.3379  4.2325 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -5.733e+03  1.232e+03  -4.654 1.16e-05 ***
t            5.924e+00  1.250e+00   4.739 8.30e-06 ***
t2          -1.530e-03  3.170e-04  -4.827 5.87e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.815 on 87 degrees of freedom
Multiple R-squared:  0.7391,	Adjusted R-squared:  0.7331 
F-statistic: 123.3 on 2 and 87 DF,  p-value: < 2.2e-16
```

Hide

Hide

```
plot(ts(fitted(model.ozone.qa)), ylim = c(min(c(fitted(model.ozone.qa),
                                                 as.vector(ozone))), max(c(
fitted(model.ozone.qa),as.vector(ozone)))),
     ylab='y' ,xlab='Years', main = "Fitted quadratic curve", type="l",lty=2,col="
red")
lines(as.vector(ozone),type="o")
```



This model has adjusted R - square value of 0.7331. The plot also gives us a slope that is statistically significant at 5% significance level.

Residual Analysis

The residuals would behave like a true stochastic component if the trend model is correct.By looking at the residual the assumptions of the stochastic component can be judged.

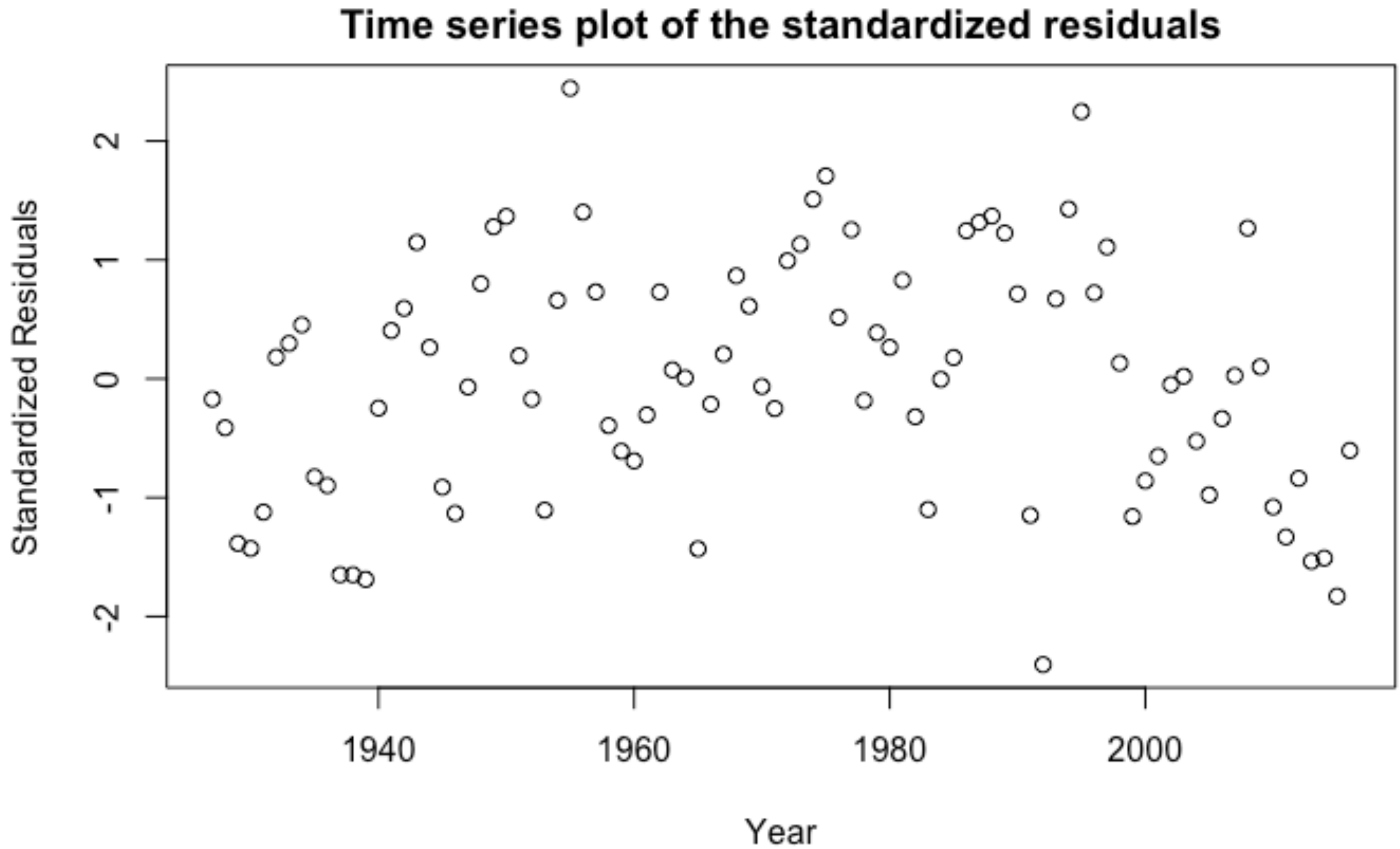
Linear Model

Scatter plot for Linear Model

Hide

Hide

```
#plotting scatter plot
plot(y=res.model.ozone.ln,x=as.vector(time(ozone)), ylab='Standardized Residuals'
  ,xlab='Year', type='p', main = "Time series plot of the standardized residuals")
```



The scatter plot for linear model depicts a slight downward trend. The plot tends to dip towards the ending years.

Histogram for linear trend model

The histogram helps to check the normality of the residuals

Hide

Hide

```
#plotting histogram
hist(res.model.ozone.ln,xlab='Standardized Residuals')
```

If the data seems to be normally distributed it would give us a symmetric graph.In the case of linear trend model the plot does seem very symmetric.

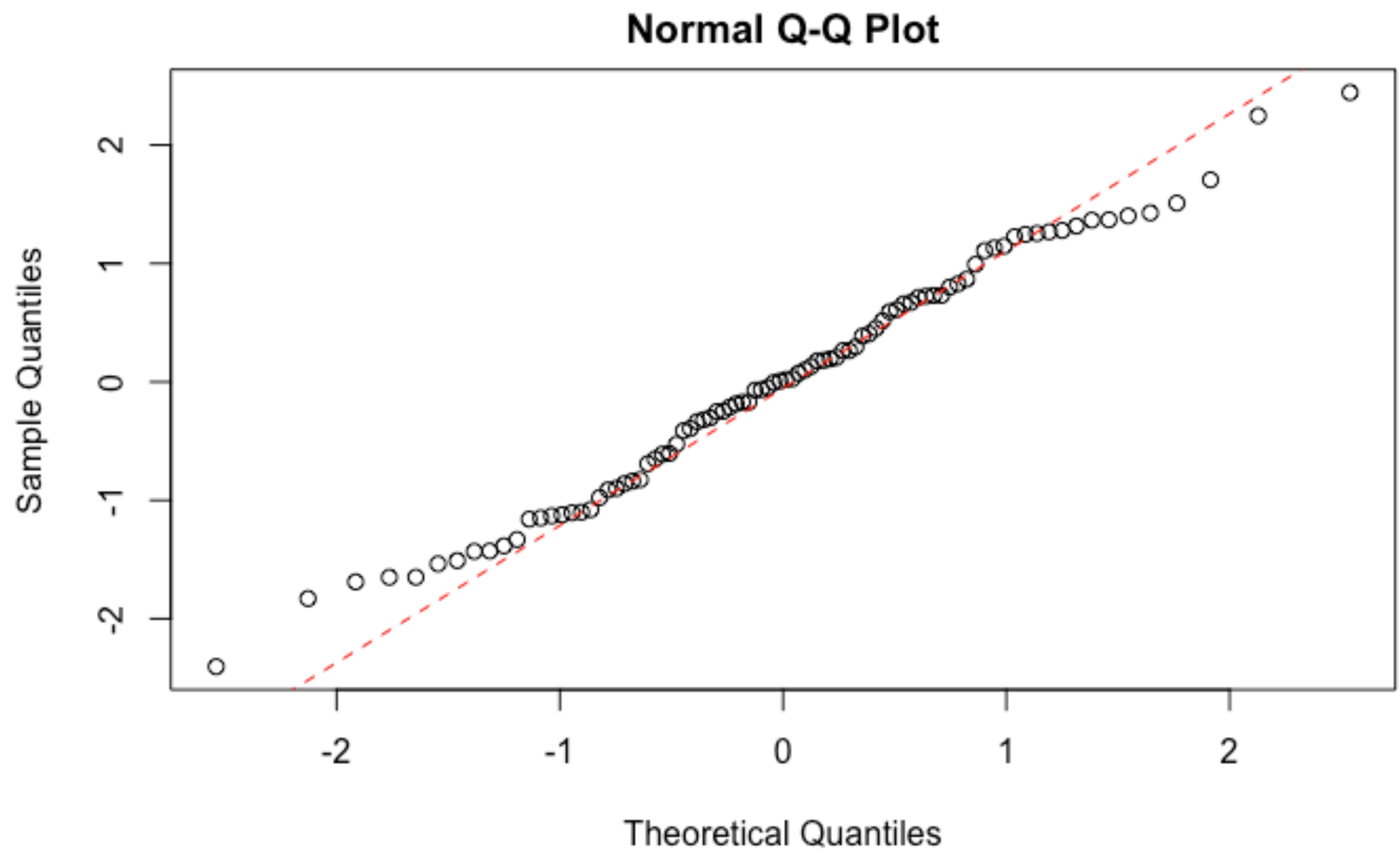
Quantile-quantile (QQ) plot for Linear Trend Model

This plot shows the quantiles of the data and the theoretical quantiles of the distribution.

Hide

Hide

```
#plotting qqplot
qqnorm(res.model.ozone.ln)
qqline(res.model.ozone.ln, col = 2, lwd = 1, lty = 2)
```



If the data is normally distributed the plot would seem to be in a straight line. In the linear trend model the line seems to be somewhat a straight line.

Shapiro-Wilk normality test for Linear Trend Model

Shapiro-Wilk normality test is a hypothesis test that is used to examine normality of the stochastic component

Hide

Hide

```
#Shapiro-Wilk normality test
shapiro.test(res.model.ozone.ln)
```

```

    Shapiro-Wilk normality test

data:  res.model.ozone.ln
W = 0.98733, p-value = 0.5372
```

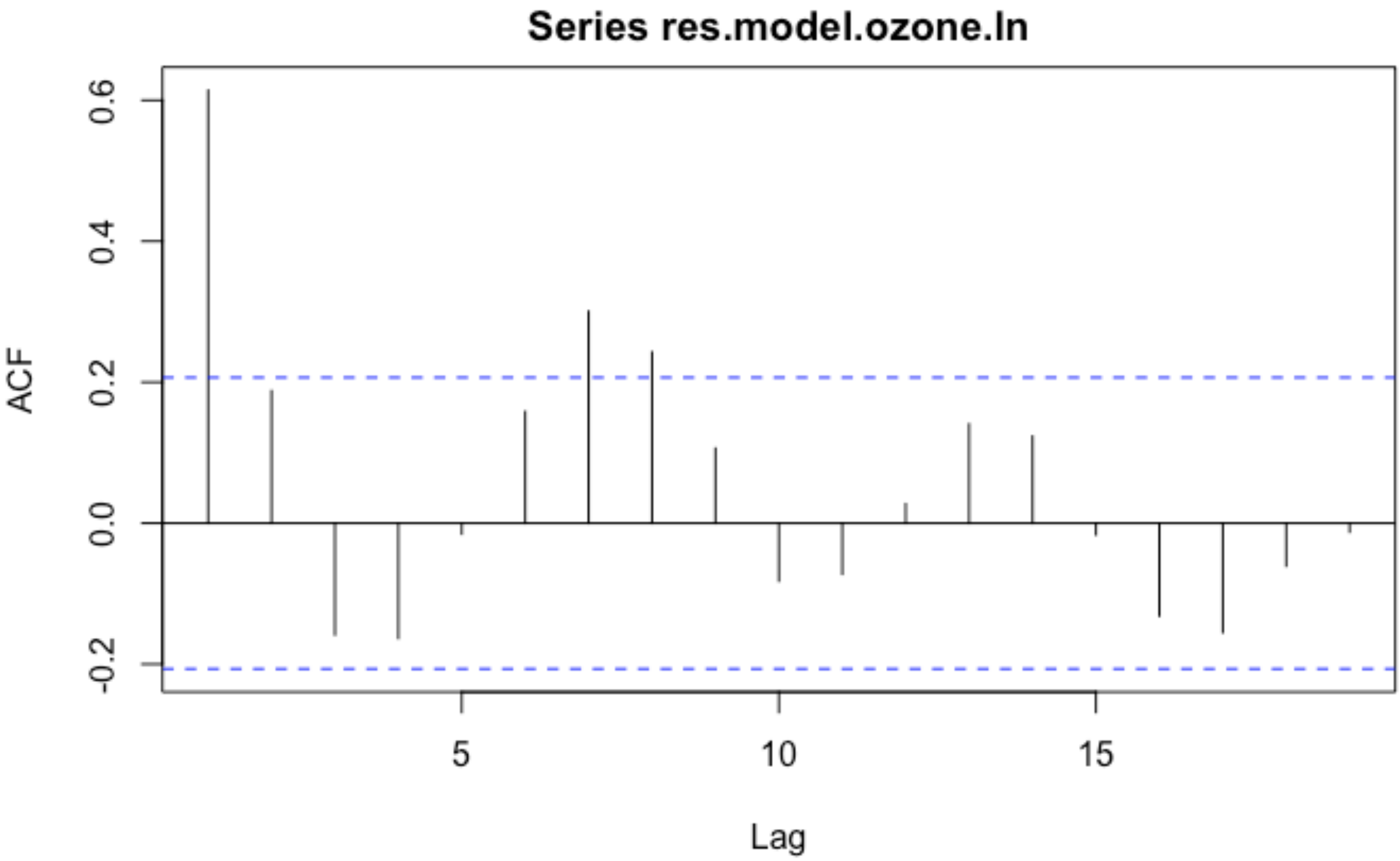
It finds the correlation between normal quantiles that corresponds to the residuals. In case of linear trend we get the p-value of 0.5372. So p-values greater than 0.05 we conclude not to reject the null hypothesis.

Sample Autocorrelation Function for Linear Trend Model

Hide

Hide

```
#plotting acf
acf(res.model.ozone.ln)
```



The smoothness of the time series plot can be seen as there are higher correlation values than confidence levels at several lags.

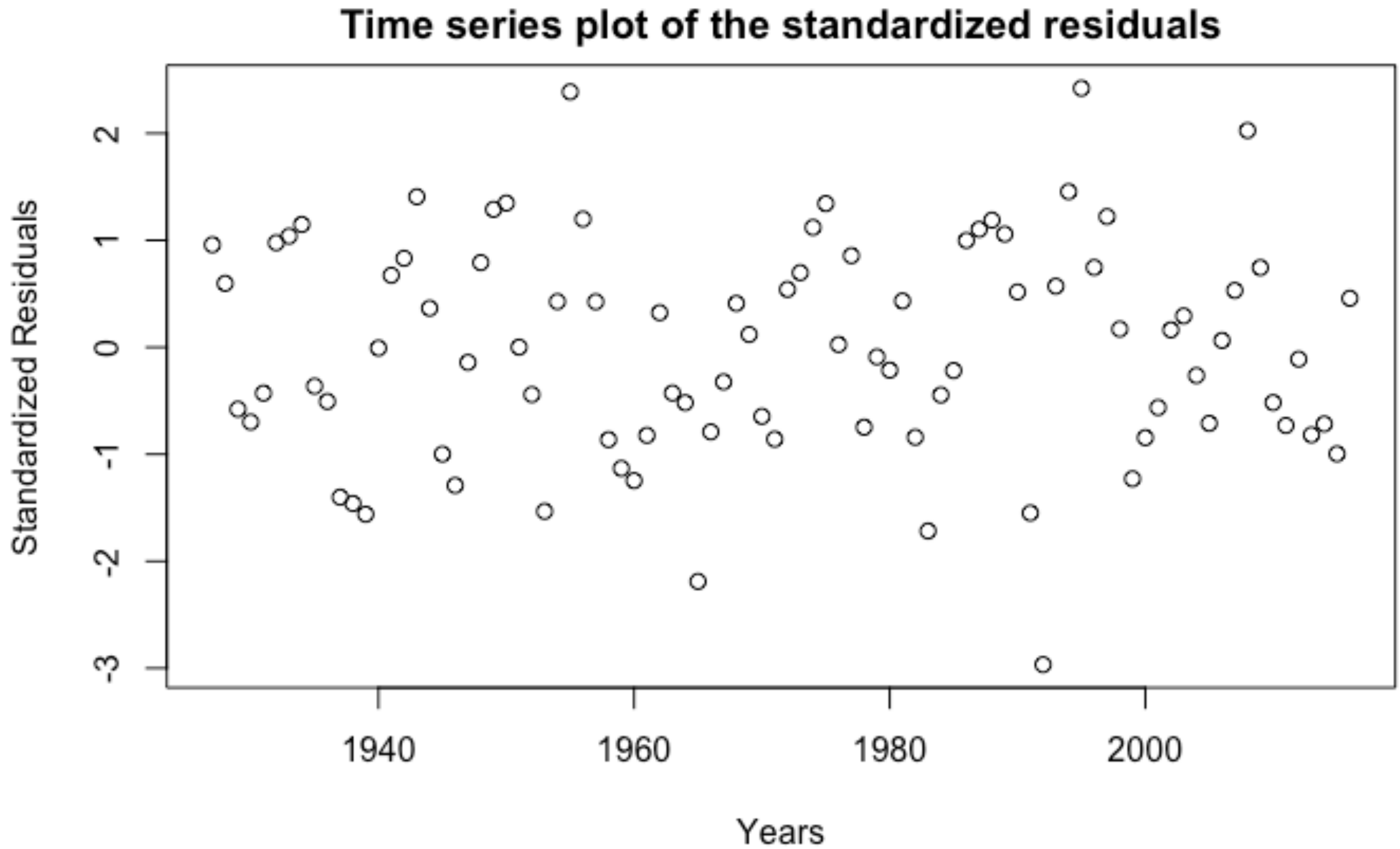
Quadractic Model

Scatter plot for Quadractic Trend Model

Hide

Hide

```
#plotting scatter plot
plot(y=res.model.ozone.qa,x=as.vector(time(ozone)), ylab='Standardized Residuals'
  ,xlab='Years', type='p', main = "Time series plot of the standardized residuals"
)
```



The scatter plot for quadractic model depicts a slight downward trend. The plot tends to dip towards the ending years.

Histogram for Quadractic Trend Model

Hide

Hide

```
#plotting histogram
hist(res.model.ozone.qa,xlab='Standardized Residuals')
```

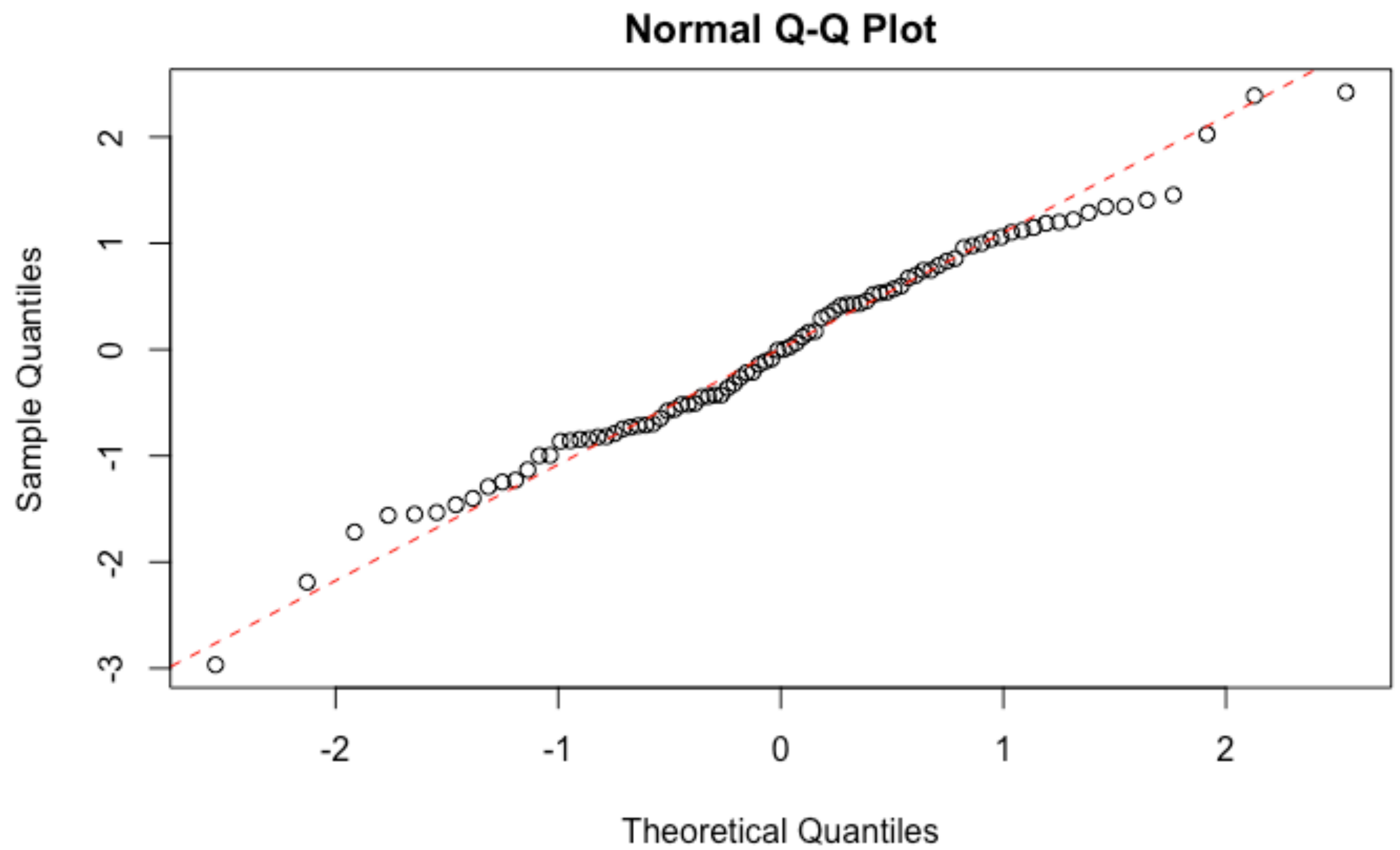
The histogram helps to check the normality of the residuals.If the data seems to be normally distributed it would give us a symmetric graph.In the case of Quadractic trend model the plot does not seem very symmetric.

Quantile-quantile (QQ) plot for Quadractic Trend Model

Hide

Hide

```
#plotting qqplot
qqnorm(res.model.ozone.qa)
qqline(res.model.ozone.qa, col = 2, lwd = 1, lty = 2)
```



If the data is normally distributed the plot would seem to be in a straight line. When compared to linear trend model the line seems to be in a better straight line in quadractic model trend thus making it normal distribution

Shapiro-Wilk normality test for Quadractic Trend Model

Shapiro-Wilk is a hypothesis test that is used to examine normality of the stochastic component.

Hide

Hide

```
# Shapiro-Wilk normality test
shapiro.test(res.model.ozone.qa)
```

```

	Shapiro-Wilk normality test

data:  res.model.ozone.qa
W = 0.98889, p-value = 0.6493
```

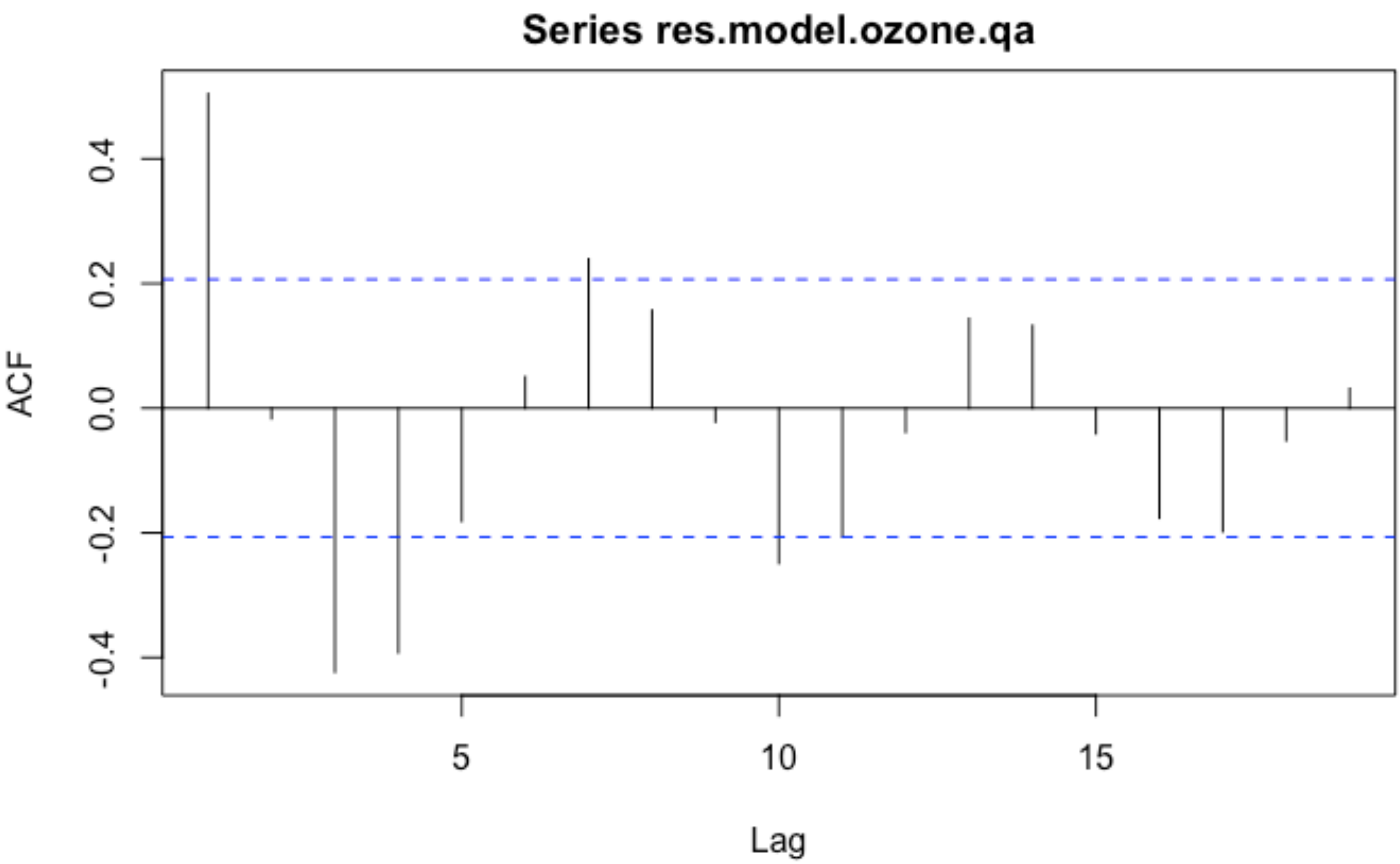
It finds the correlation between normal quantiles that corresponds to the residuals. In case of Quadractic trend we get the p-value of 0.6493, so p-values greater than 0.05 we conclude not to reject the null hypothesis.

Sample Autocorrelation Function for Quadractic Trend Model

Hide

Hide

```
#plotting acf
acf(res.model.ozone.qa)
```



The smoothness of the time series plot can be seen as there are higher correlation values than confidence levels at several lags.

Forecasting with regression models

Hide

Hide

```
# Create time points for model fitting
t =c(2017,2018,2019,2020,2021)
t2 = t^2
new = data.frame(t,t2)
forecasts =predict(model.ozone.qa, new, interval ="prediction")
print(forecasts)
```

```
        fit       lwr       upr
1 -10.34387 -14.13556 -6.552180
2 -10.59469 -14.40282 -6.786548
3 -10.84856 -14.67434 -7.022786
4 -11.10550 -14.95015 -7.260851
5 -11.36550 -15.23030 -7.500701
```

Hide

Hide

```
#plotting the forecasting
plot(ozone, xlim = c(1927,2022), ylim =c(-15, 10), ylab ="Ozone Layer",xlab="Year"
,main="Ozone Layer Thickness for next 5 years")
#forecast converted to time series
op <- par(cex=0.9)
```

Hide

Hide

```
lines(ts(as.vector(forecasts[,1]), start =2017), col="red", type="l",lwd=2)
lines(ts(as.vector(forecasts[,2]), start =2017), col="blue", type="l",lwd=2)
```

Hide

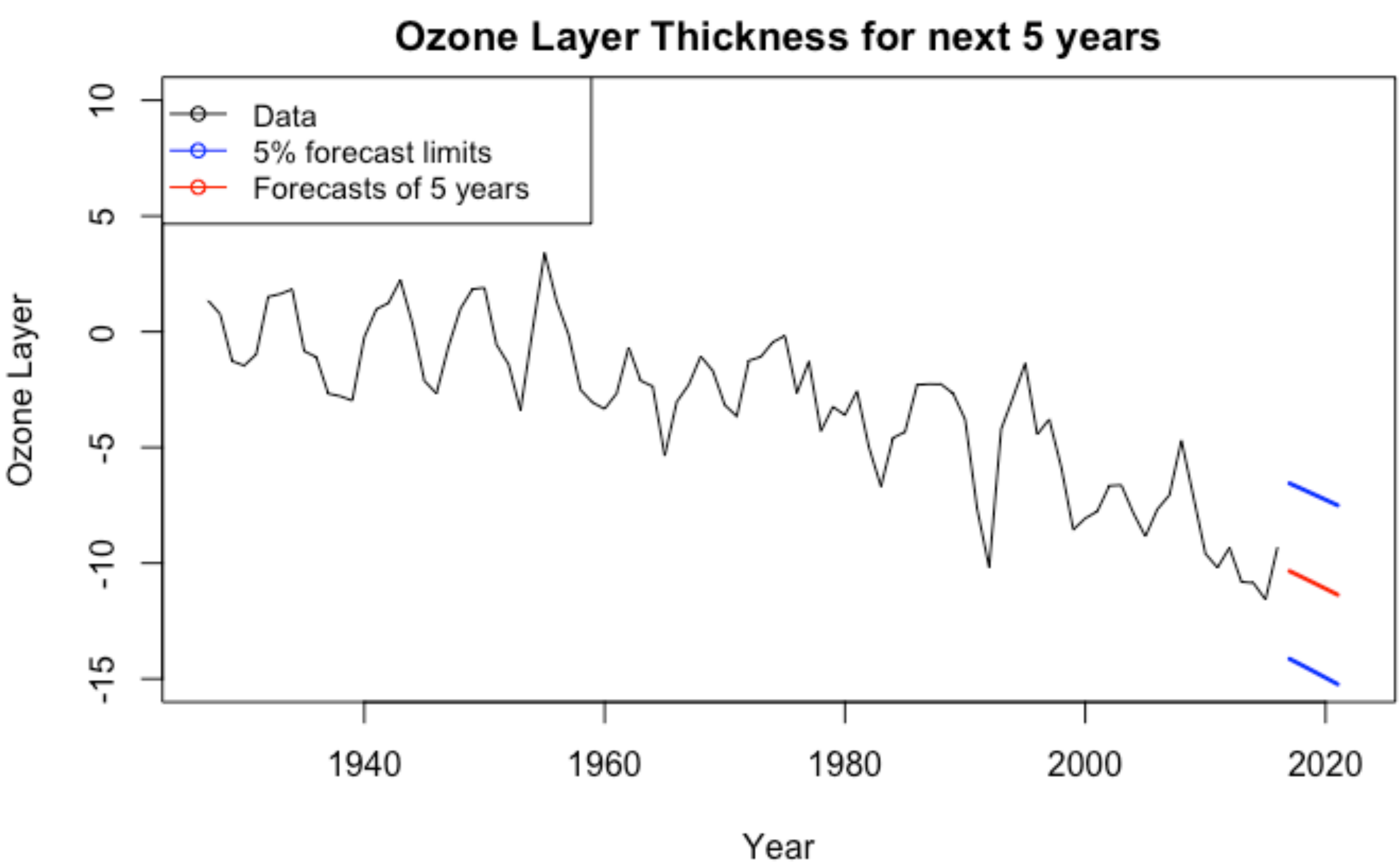
Hide

```
lines(ts(as.vector(forecasts[,3]), start =2017), col="blue", type="l",lwd=2)
legend("topleft", lty=1, pch=1, col=c("black","blue","red"), text.width =27,
       c("Data","5% forecast limits", "Forecasts of 5 years"))
```

Hide

Hide

```
par(op)
```



The ozone thickness of the coming 5 years were predicted by forecasting. From the residual analysis we were able to choose quadractic model over the linear model.Forecating was later done on the quadractic trend model. The plot depicts a repeated pattern that predicts a downward trend within the next 5 years. Thus the ozone layer thickness would decrease over the years.

Conclusion

With the ozone thickess data that was provided to us we were able to conclude that even though we had a good linear model the quadractic model performs better.There is an added variance of 6% in quadractic model and according to the qqplot that was plotted during residual analysis the quadractic plot tends to give us a straighter line compared to linear trend. When comparing the Shapiro-Wilk normality test the Quadractic trend tends to have a larger value even by points, proving that larger W values indicate your sample is normally distributed. Once the quadractic model was choosen forecasting was done to predict the thickness of ozone layer wthin 5 years,where we were able to come to the conclusion that it would decrease over the years.

CHAPTER 2

Introduction

In task 2 we again plot the time series for the ozone layer thickness dataset.ACF and PACF was used to notice the patterns within the plot. Differencing and transformation was done on the series if it was found to have changing variance. This helps the series to be stationary and homogeneously distributed. These plots also help to determine ARIMA(p,d,q).The orders of ARIMA(p,d,q) model is done using ACF, PACF and EACF plots and BIC table.

ACF and PACF

Hide

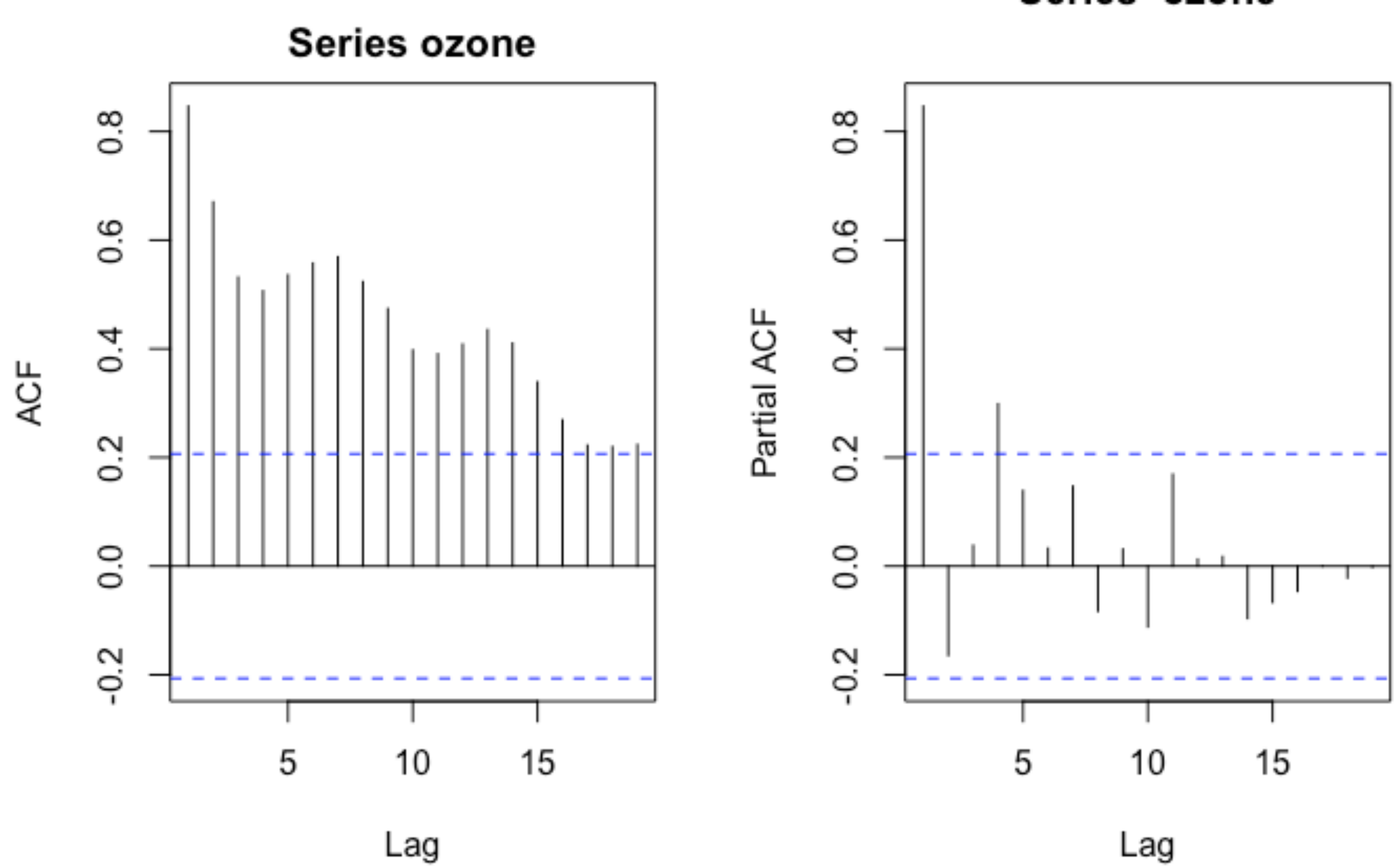
Hide

```
 # To plot both ACF and PACF
par(mfrow=c(1,2))
acf(ozone)
pacf(ozone)
```

Hide

Hide

```
par(mfrow=c(1,1))
```



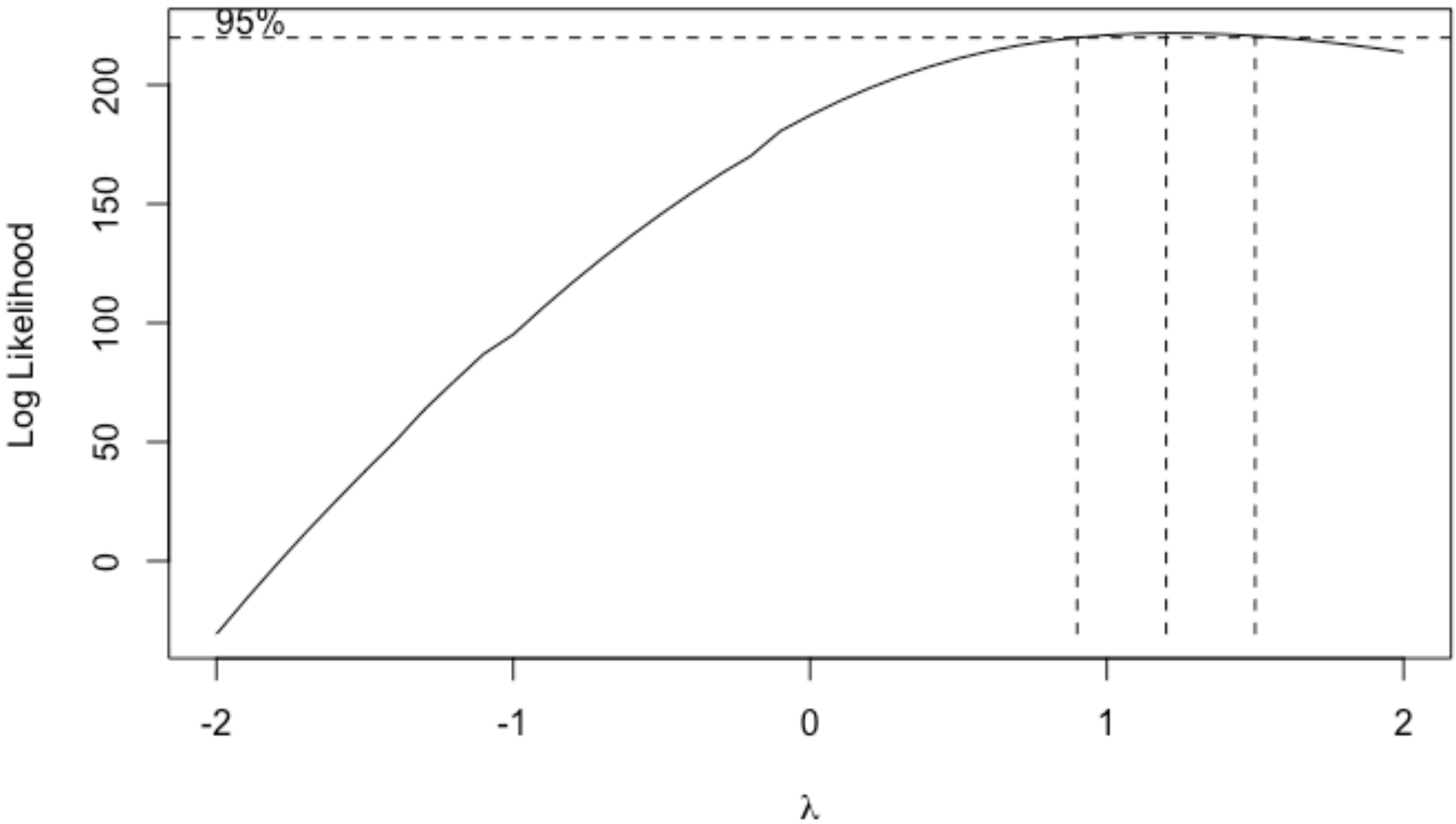
ACF has a slow decaying pattern while PACF has its first correlation as a very high value. This implies that there is existence of nonstationarity and trend.

BOX-COX Plot

Hide

Hide

```
#plotting box-cox
ozone.transform = BoxCox.ar(ozone + abs(min(ozone))+1)
```



Hide

Hide

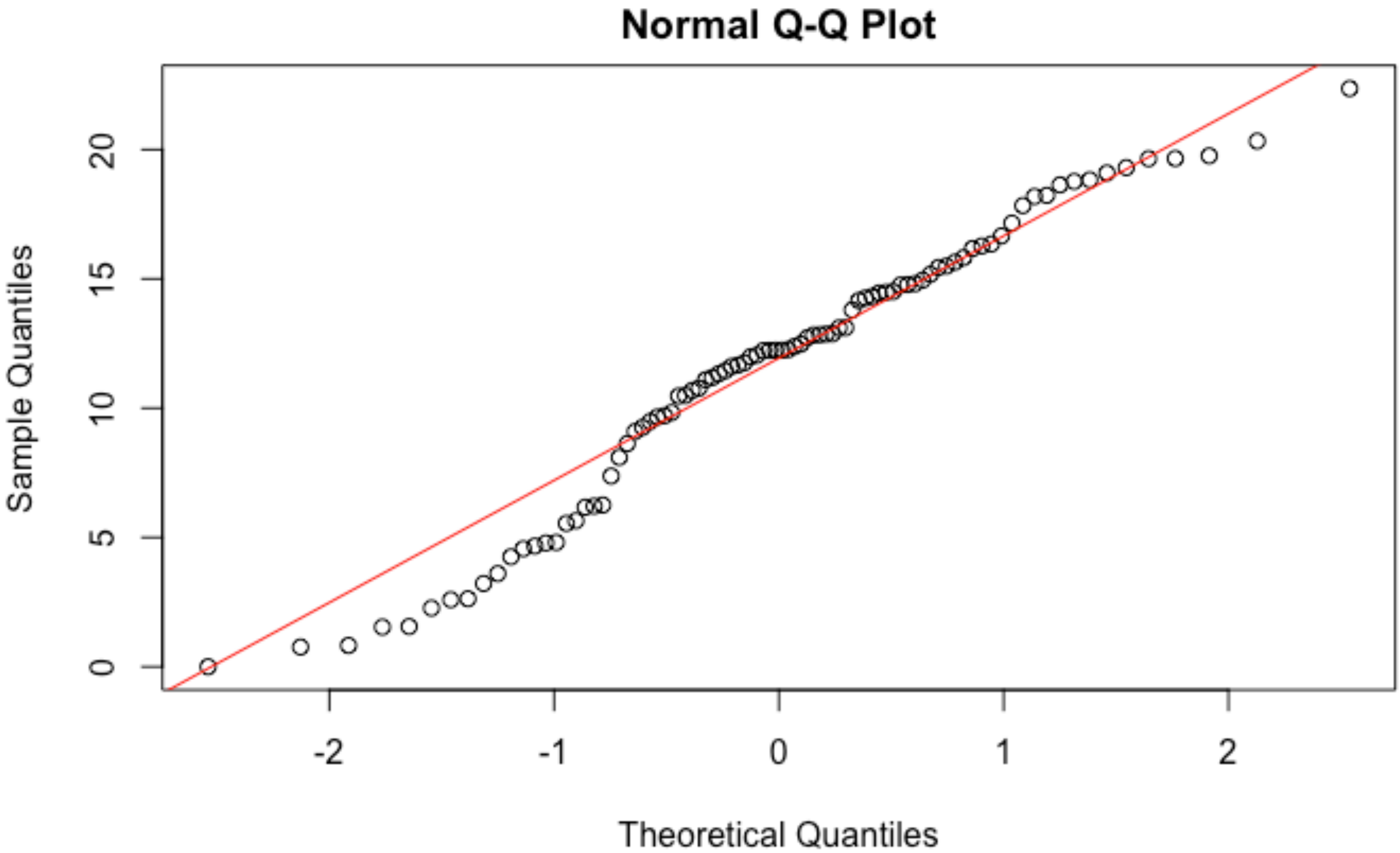
```
#lambda value
ozone.transform$ci
```

```
[1] 0.9 1.5
```

Hide

Hide

```
lambda = 1.2
ozone = ozone + abs(min(ozone))+1
BC.ozone = (ozone^lambda-1)/lambda
#plotting qqplot
qqnorm(BC.ozone)
qqline(BC.ozone, col = 2)
```



Hide

Hide

```
#Shapiro-Wilk normality test
shapiro.test(BC.ozone)
```

```

    Shapiro-Wilk normality test

data:  BC.ozone
W = 0.96644, p-value = 0.01995
```

The confidence interval 95% for contains the values of between 0.9 and nearly 1.8 with the center of confidence interval corresponds to approximately 1.2.

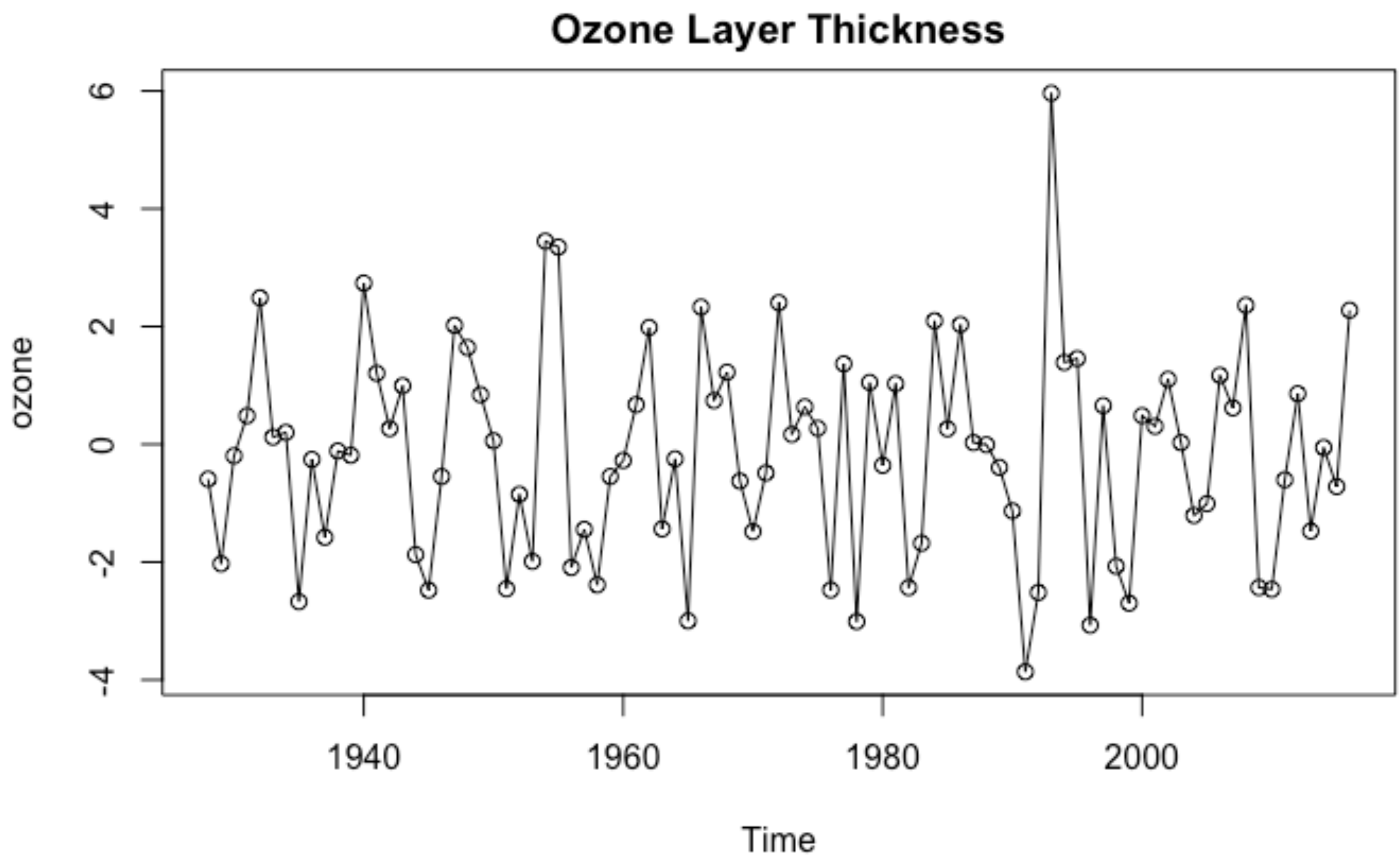
Differencing the Data

The interval has a mid point closer to 1 which concluded to no transformation. Decided to move forward with Differencing.

Hide

Hide

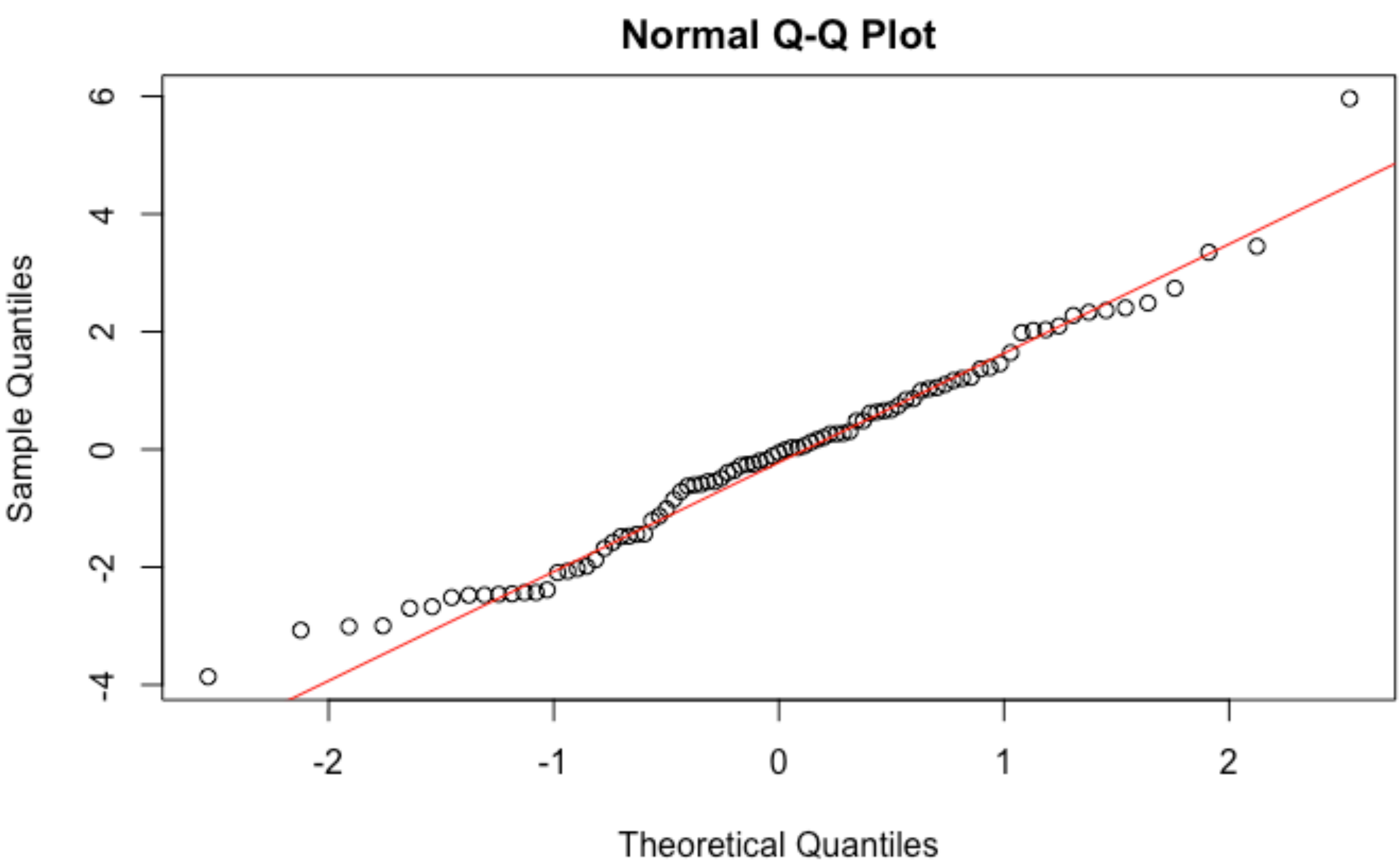
```
#differencing the original data
diff.ozone = diff(ozone)
#plotting time series plot
plot(diff.ozone,type='o',ylab='ozone',main="Ozone Layer Thickness")
```



Hide

Hide

```
#plotting qqplot
qqnorm(diff.ozone)
qqline(diff.ozone, col = 2)
```



After differencing the original data the qqplot provides us with a straighter line when compared to the plot before differencing.

Shapiro-Wilk normality test

Hide

Hide

```
# Shapiro-Wilk normality test
shapiro.test(diff.ozone)
```

```
    Shapiro-Wilk normality test

data:  diff.ozone
W = 0.97907, p-value = 0.1606
```

In case of Shapiro-Wilk normality test we get the p-value of 0.1606. So p-values greater than 0.05 we conclude not to reject the null hypothesis.

Hide

Hide

```
# Augmented Dickey-Fuller Test
adf.test(diff.ozone)
```

```
    Augmented Dickey-Fuller Test

data:  diff.ozone
Dickey-Fuller = -7.1568, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary
```

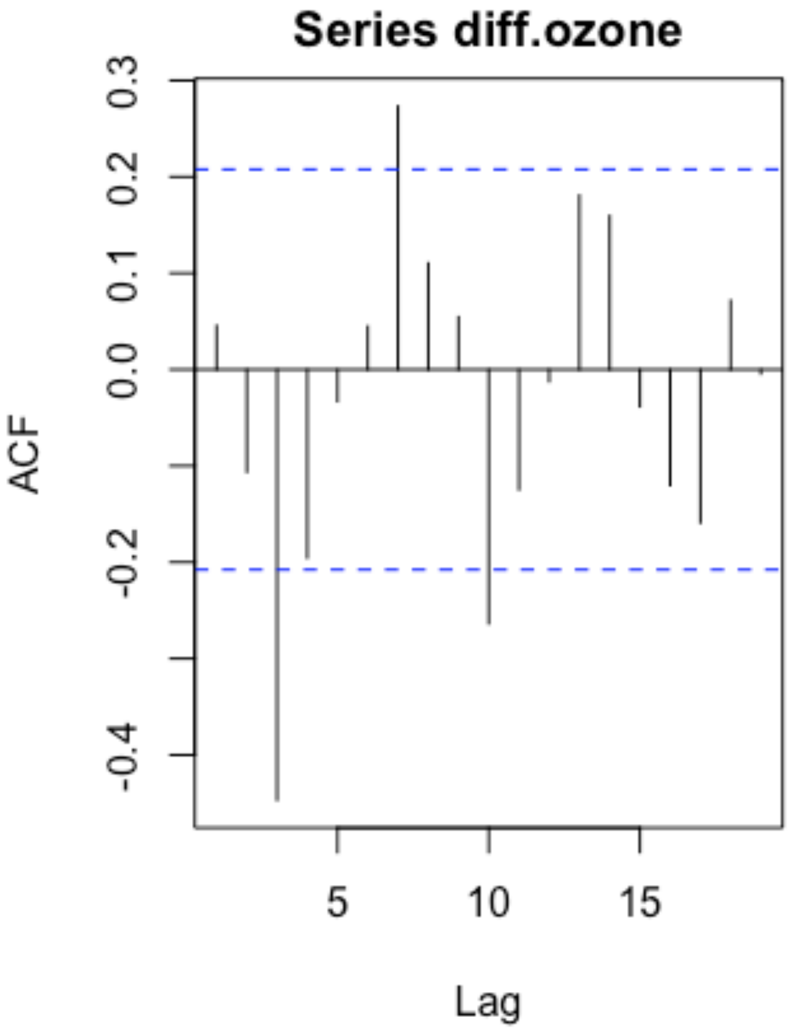
The p-value in Augmented Dickey-Fuller Test is 0.01 and the alternative hypothesis is said to be stationary with the first differencing.

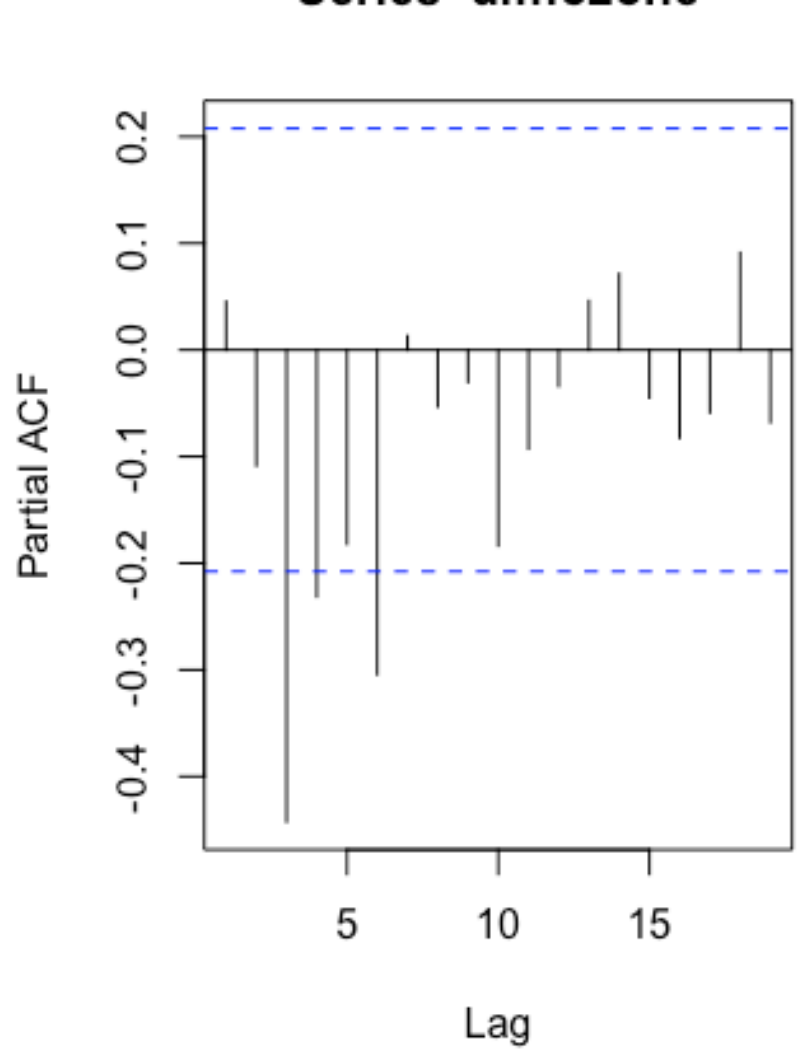
ACF and PACF

Hide

Hide

```
# To plot both ACF and PACF after differencing
par(mfrow=c(1,2))
#plotting ACF
acf(diff.ozone)
#plotting PACF
pacf(diff.ozone)
```





There is three significant lag in ACF and two significant lags in PACF. From this plot we can depict MA(3), AR(2) and ARIMA(3,1,2)

EACF plot

Hide

Hide

```
#plotting EACF
eacf(diff.ozone)
```

```
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 o o x o o o x o o x o  o  o  o
1 x o x o o o o o o x o  o  o  o
2 x o x o o o x o o x o  o  o  o
3 x o x o o x o o o o o  o  o  o
4 x o o x o x o o o o o  o  o  o
5 x x x x o x o o o o o  o  o  o
6 o o o x x o o o o o o  o  o  o
7 o o o x o o o o o o o  o  o  o
```

The eacf help us to identify the triangle providing us ARIMA(3,1,3) and ARIMA(3,1,1) models as the other combinations give large models.

BIC table

Hide

Hide

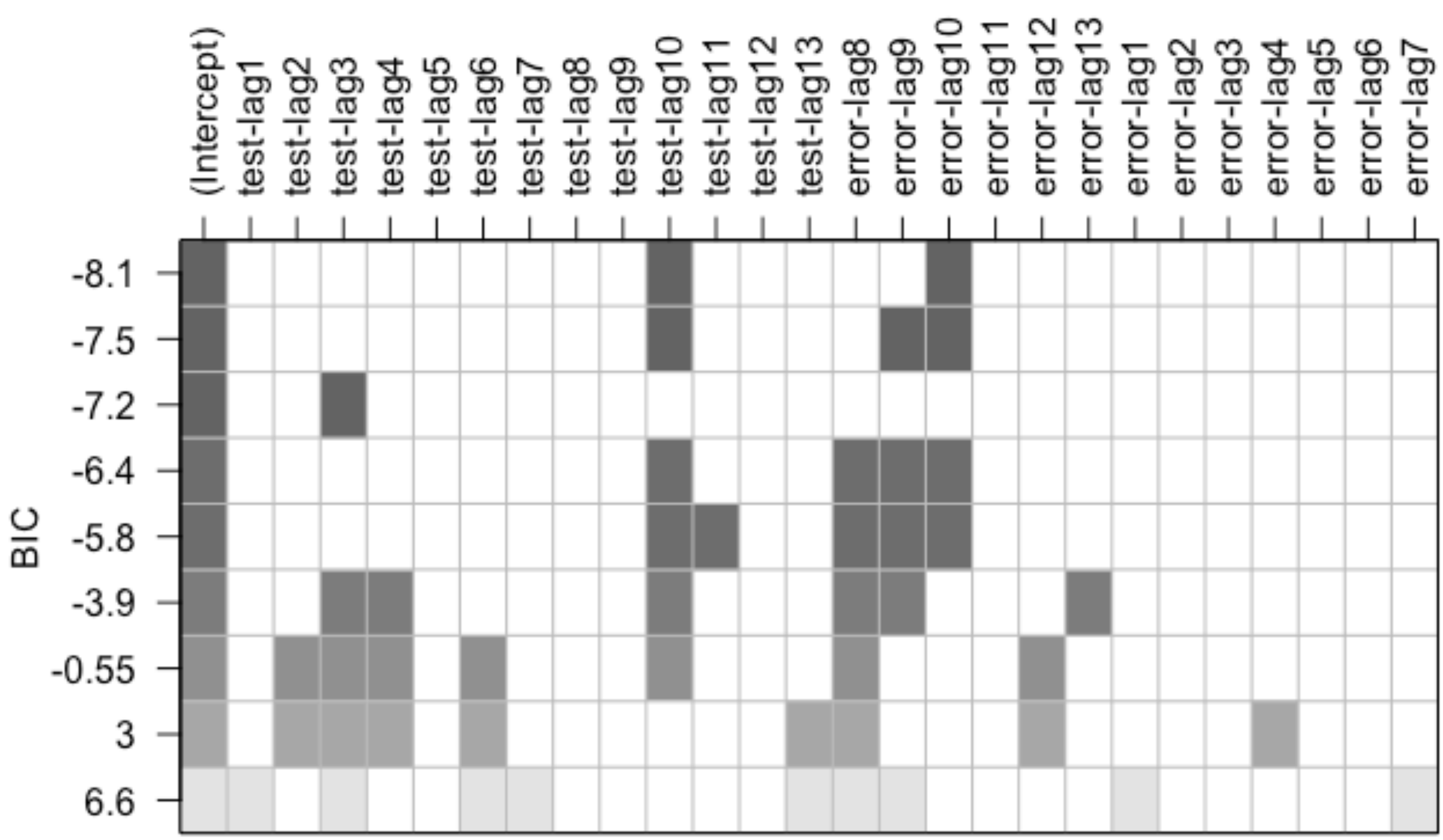
```
par(mfrow=c(1,1))
#plotting BIC Table
bic_ozone = armasubsets(y=diff.ozone,nar=13,nma=13,y.name='test',ar.method='ols')
```

```
Reordering variables and trying again:
```

Hide

Hide

```
plot(bic_ozone)
```



In the BIC table, we have obtained a darker shaded test-lag10, but because MA(10) is a higher value it is discarded and the MA(3) coefficients is considered.Thus making the models ARIMA(3,1,2), ARIMA(3,1,3) and ARIMA(3,1,1).

Conclusion

In conclusion,as there was change in variance found in the dataset differencing was done. Thus making the dataset stationary and homogeneously distributed. Using ACF, PACF AND EACF plots provide us with possible models such as ARIMA(3,1,2), ARIMA(3,1,3) and ARIMA(3,1,1).