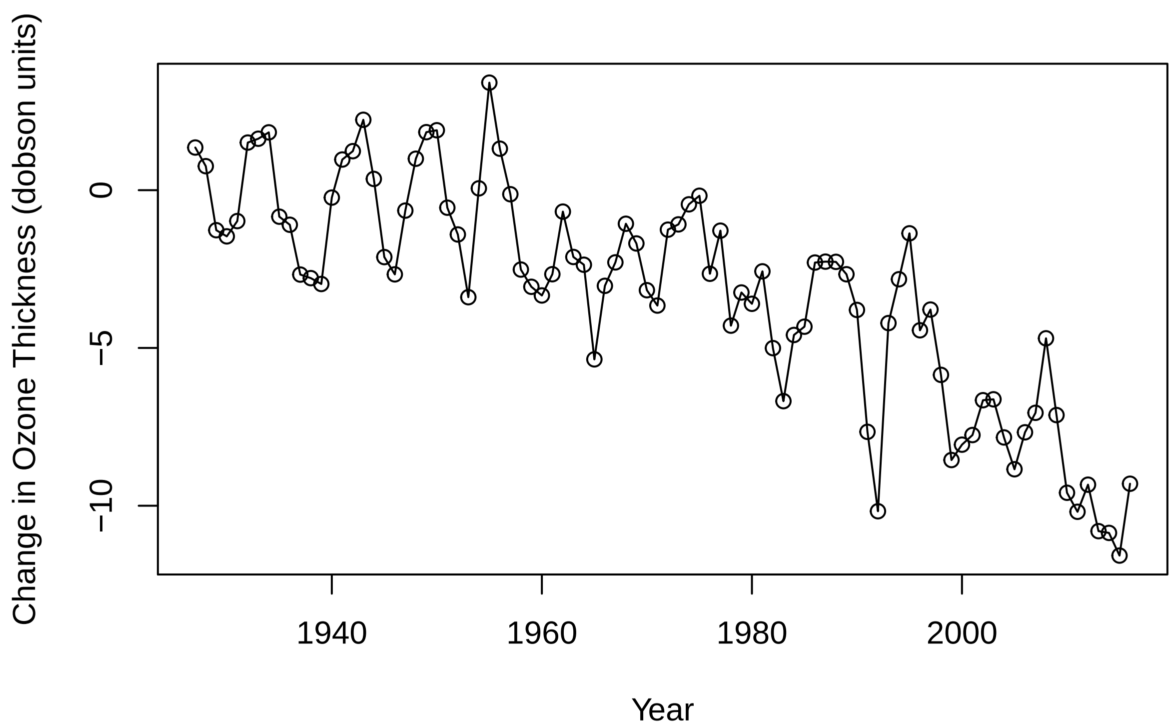
# MATH1318 Time Series Analysis Assignment 1

*Ashleigh Olney s3686808*

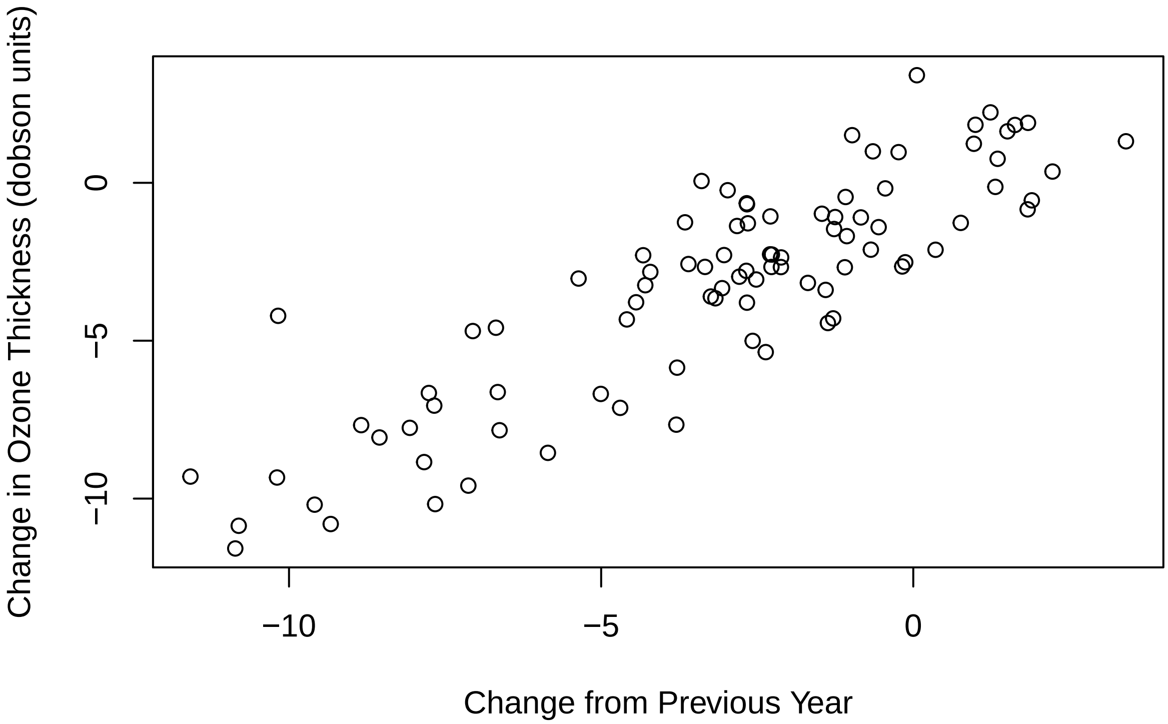
*19/03/2019*

## Data

```
# import data
data1 <- read.csv("data1.csv", header = FALSE)
rownames(data1) <- seq(from=1927, to=2016)

# convert to ts object
data1 <- ts(as.vector(data1), start=1927, end=2016)

# generate basic plot
plot(data1, type = 'o', xlab = "Year", ylab = "Change in Ozone Thickness (dobson units)")
```




```
# investigate correlation with previous year change
plot(y = data1,
     x = zlag(data1),
     ylab = "Change in Ozone Thickness (dobson units)",
     xlab = "Change from Previous Year")
```



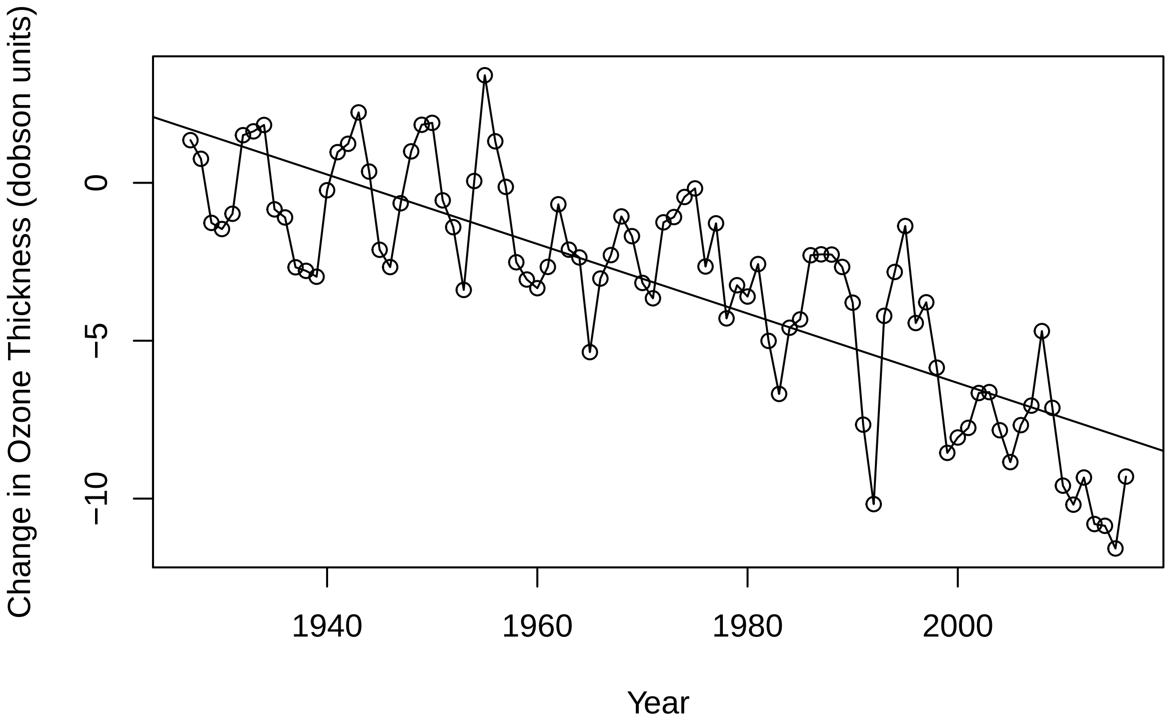

```
# generate R value for previous year change
x = zlag(data1)
index = 2:length(x)
cor(data1[index],x[index])
```

```
## [1] 0.8700381
```

Correlation of change in ozone and the previous year's change is strong, 87% of the variation is explained by this correlation.

## Linear Model

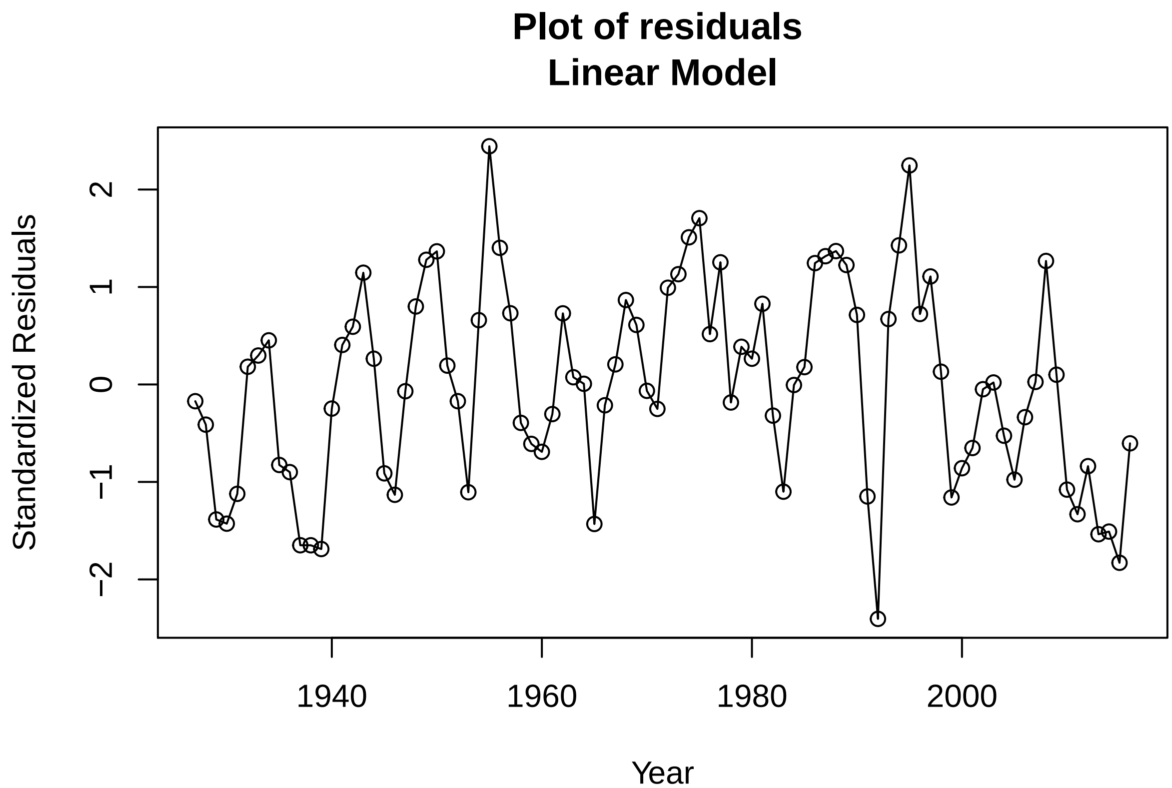
```
# generate basic plot
plot(data1, type = 'o', xlab = "Year", ylab = "Change in Ozone Thickness (dobson units)")

# linear model
t = time(data1)
model1 = lm(data1 ~ t)

# add linear model to plot
abline(model1)
```



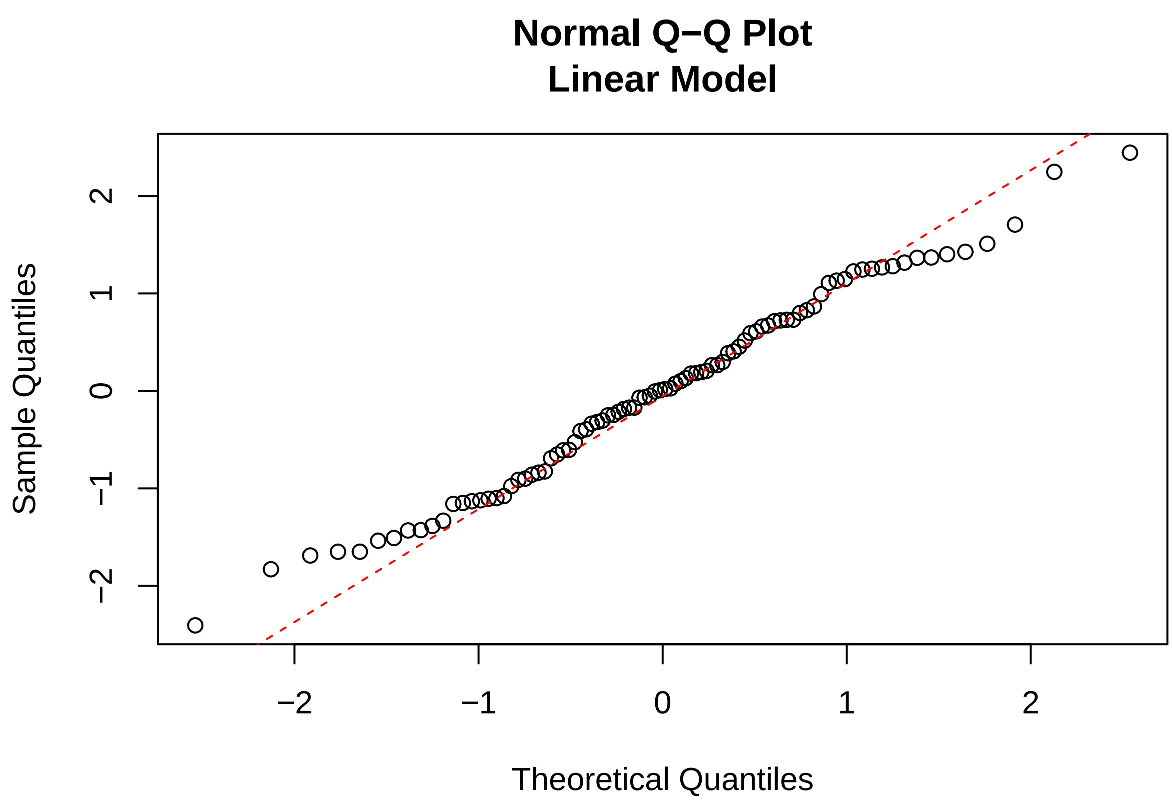

```
summary(model1)
```

```
##
## Call:
## lm(formula = data1 ~ t)
##
## 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 ***
## t            -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
```

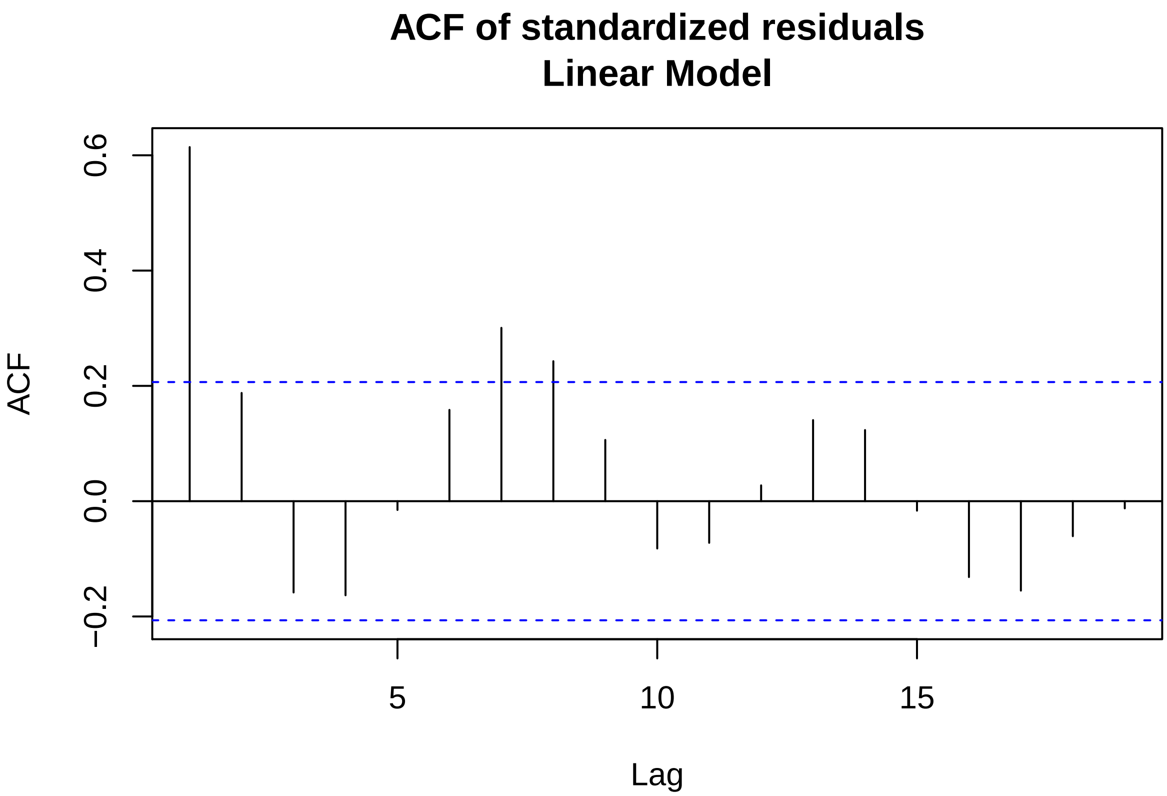
```
# plot residuals - linear model
plot(y=rstudent(model1),x=as.vector(time(data1)),
     xlab='Year',ylab='Standardized Residuals',
     type='o', main = "Plot of residuals \nLinear Model")
```




```r
# qqplot - linear model
y = rstudent(model1)
qqnorm(y, main = "Normal Q-Q Plot\nLinear Model")
qqline(y, col = 2, lwd = 1, lty = 2)
```



```r
# acf - linear model
acf(rstudent(model1), main = "ACF of standardized residuals\nLinear Model")
```




```r
# normality test - linear model
shapiro.test(rstudent(model1))
```

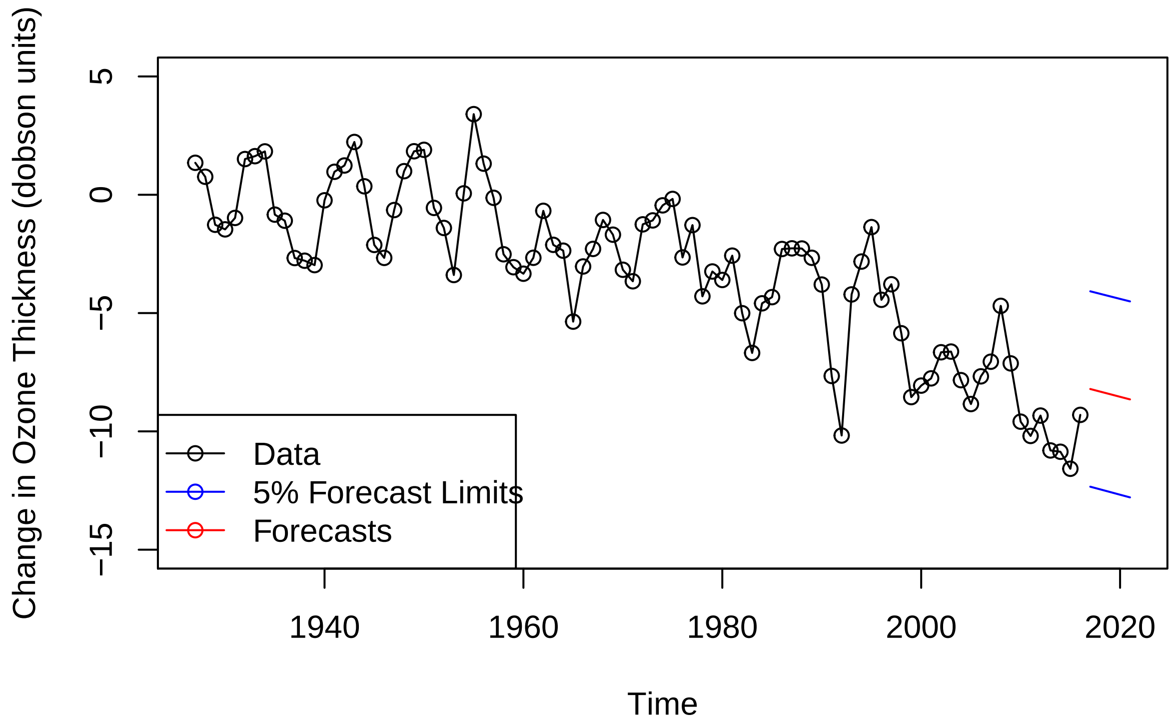
```
##
##  Shapiro-Wilk normality test
##
## data:  rstudent(model1)
## W = 0.98733, p-value = 0.5372
```

```r
# forecast linear model
newdata = data.frame(t = c(2017:2021))
forecastLinear = predict(model1, newdata, interval = "prediction")
print(forecastLinear)
```

```
##         fit       lwr       upr
## 1 -8.208590 -12.33732 -4.079864
## 2 -8.318619 -12.45033 -4.186902
## 3 -8.428648 -12.56342 -4.293878
## 4 -8.538677 -12.67656 -4.400792
## 5 -8.648706 -12.78977 -4.507643
```

```r
plot(data1, xlim = c(1927,2021),
     ylim = c(-15, 5),
     ylab = "Change in Ozone Thickness (dobson units)",
     type = 'o')
lines(ts(as.vector(forecastLinear[,1]), start = 2017), col="red", type="l")
lines(ts(as.vector(forecastLinear[,2]), start = 2017), col="blue", type="l")
```

```r
lines(ts(as.vector(forecastLinear[,3]), start = 2017), col="blue", type="l")
legend("bottomleft",
       lty=1, pch=1,
       col=c("black","blue","red"),
       text.width = 25,
       c("Data","5% Forecast Limits", "Forecasts"))
```

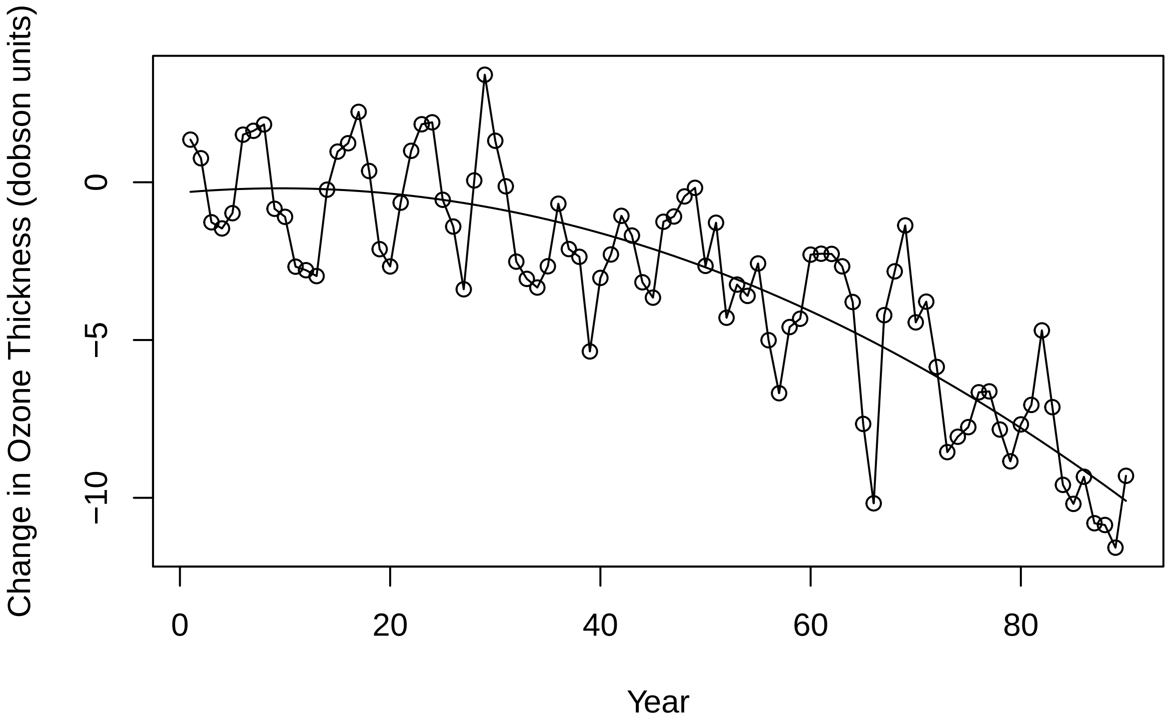



## Interpretation of Linear Model

- $\beta 0 = 213.70, \beta 1 = -0.11$
- $\mu(t) = 213.70 - 0.11(t)$
- Coefficients are significant (p $<$ 0.05)
- Model is significant (p $<$ 0.05)
- Adjusted R-squared is 0.6655, which is not strong (66% of the variation is explained by this model)
- Plot of residuals is relatively smooth.
- Q-Q Plot is not along one line, therefor does not support assumption of normality
- ACF plot has several correlation values outside of the confidence region, therefor the stochastic component of the series is not white noise. ACF plot reduces exponentially so time series may be stationary.
- Shapiro Wilks p-value of 0.5372, therefor fail to reject the null hypothesis that the stochastic component of this model is normally distributed.

## Quadratic Model

```r
# quadratic model
t = time(data1)
t2 = t^2
model2 = lm(data1~t+t2)

# plot quadratic model
plot(ts(fitted(model2)),
  ylim = c(min(c(fitted(model2),
  as.vector(data1))),
  max(c(fitted(model2),
  as.vector(data1)))),
  xlab = "Year",
  ylab = "Change in Ozone Thickness (dobson units)")
lines(as.vector(data1),type="o")
```

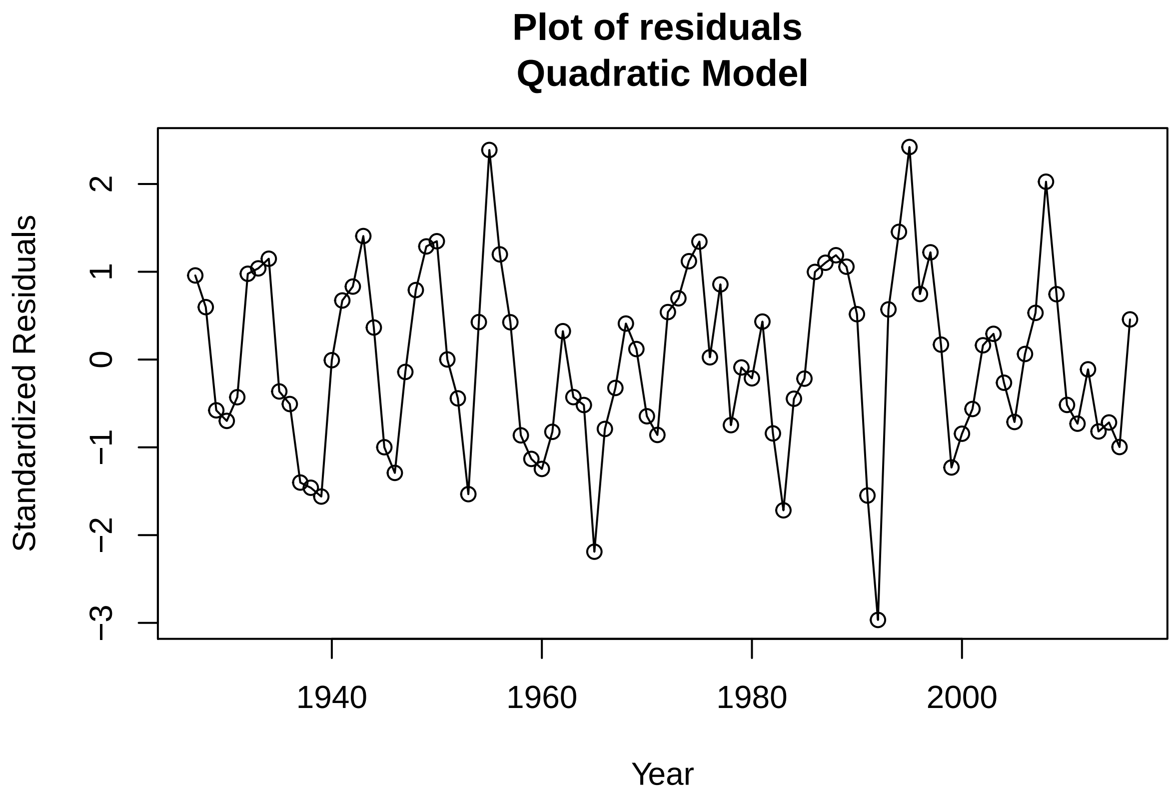



```r
summary(model2)
```

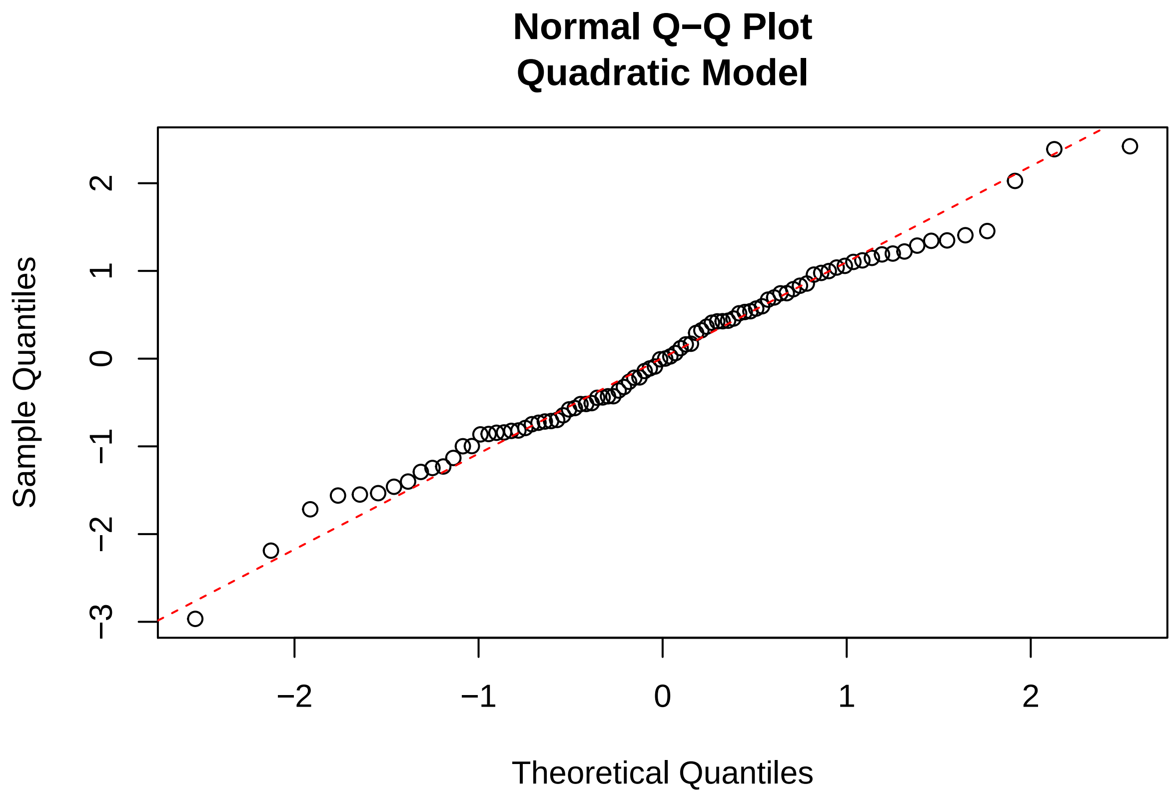
```
##
## Call:
## lm(formula = data1 ~ t + t2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.1062 -1.2846 -0.0055  1.3379  4.2325
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5.733e+03  1.232e+03  -4.654 1.16e-05 ***
## t            5.924e+00  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
```

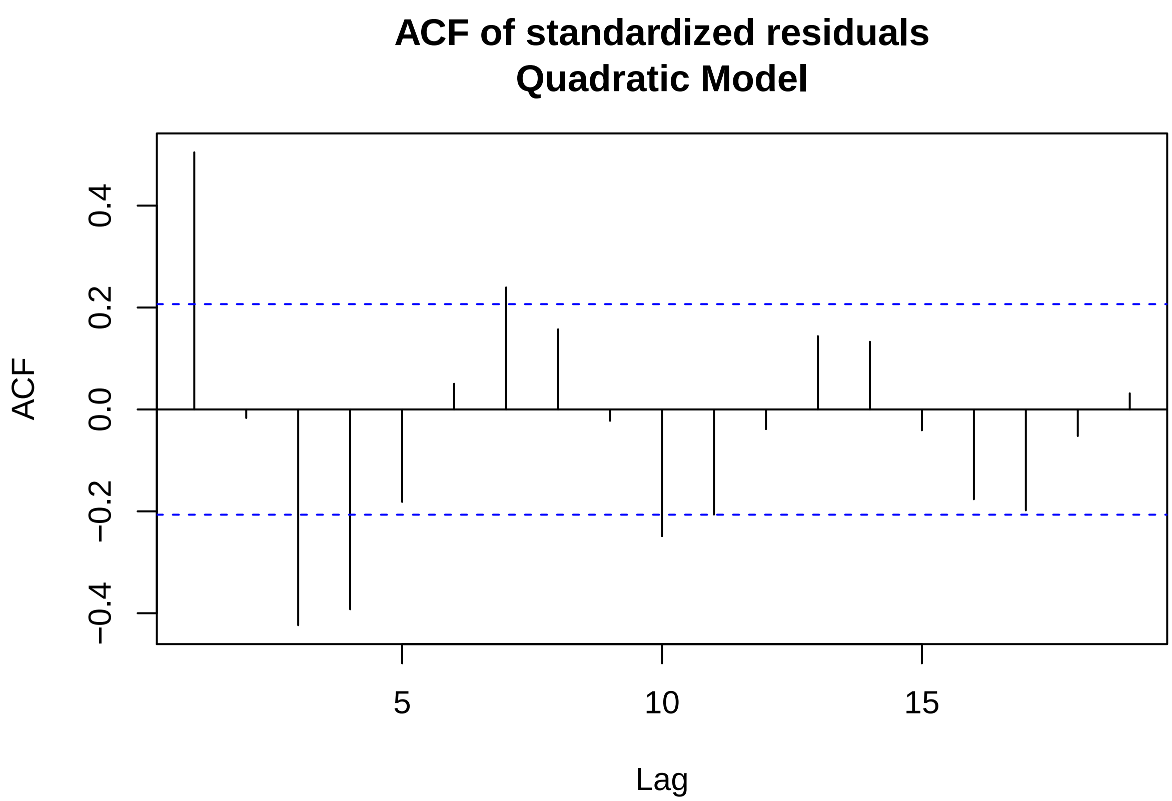
```r
# plot residuals - quadratic model
plot(y = rstudent(model2),
     x = as.vector(time(data1)),
  xlab ='Year',ylab ='Standardized Residuals',
  type ='o', main = "Plot of residuals \nQuadratic Model")
```




```r
# qqplot - quadratic model
y = rstudent(model2)
qqnorm(y, main = "Normal Q-Q Plot\nQuadratic Model")
qqline(y, col = 2, lwd = 1, lty = 2)
```




```r
# acf - quadratic model
acf(rstudent(model2), main = "ACF of standardized residuals\nQuadratic Model")
```



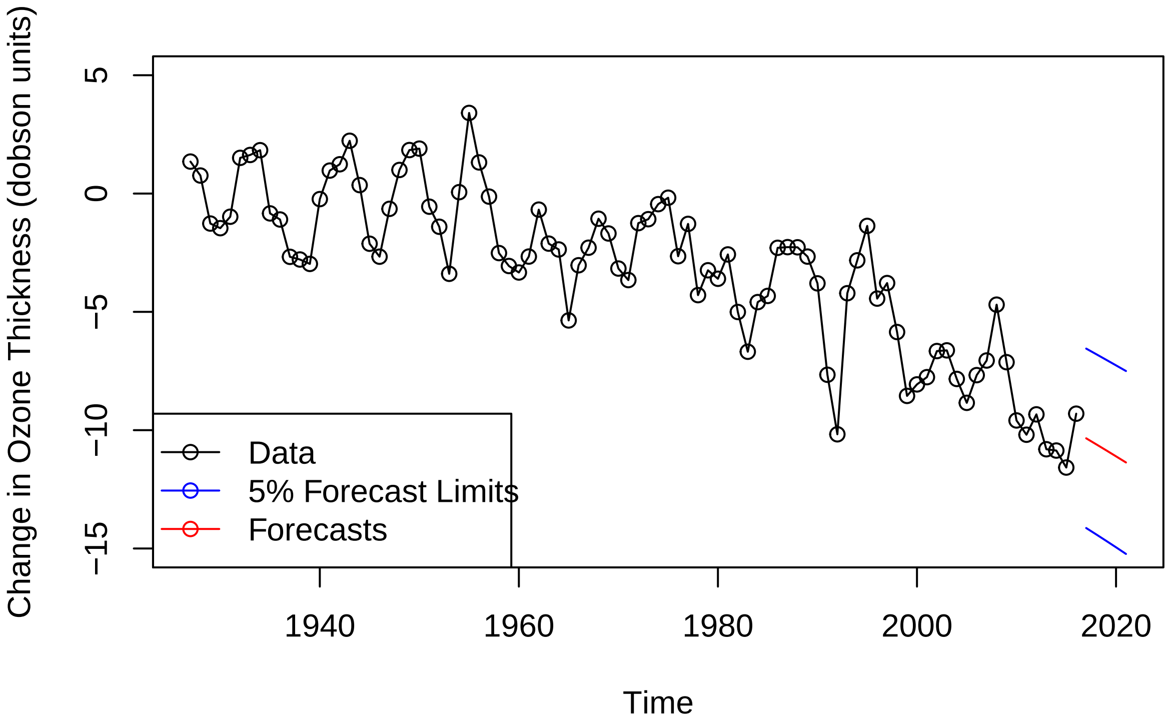
```r
# normality test - quadratic model
shapiro.test(rstudent(model2))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  rstudent(model2)
## W = 0.98889, p-value = 0.6493
```

```r
# forcast quadratic model
t = c(2017, 2018, 2019, 2020, 2021)
t2 = t^2
newdata1 = data.frame(t,t2)
forecastQuadratic = predict(model2, newdata1, interval = "prediction")
print(forecastQuadratic)
```

```
##         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
```

```r
plot(data1, xlim = c(1927,2021),
     ylim = c(-15, 5),
     ylab = "Change in Ozone Thickness (dobson units)",
     type = 'o')
lines(ts(as.vector(forecastQuadratic[,1]), start = 2017), col="red", type="l")
lines(ts(as.vector(forecastQuadratic[,2]), start = 2017), col="blue", type="l")
lines(ts(as.vector(forecastQuadratic[,3]), start = 2017), col="blue", type="l")
legend("bottomleft", lty=1, pch=1,
       col=c("black","blue","red"),
       text.width = 25 ,c("Data","5% Forecast Limits",
                          "Forecasts"))
```




## Interpretation of Quadratic Model

- $\beta 0 = -5.733e + 03, \beta 1 = -5.924, \beta 2 = -1.530e - 03$
- $\mu(t) = -5.733e + 03 - 5.924t - 1.530e - 03t^2$
- Coefficients are significant ($p < 0.05$)
- Model is significant ($p < 0.05$)
- Adjusted R-squared is 0.7331, which is moderate (73% of the variation is explained by this model).
- Plot of residuals is relatively smooth.
- Q-Q Plot is not along one line, therefor does not support assumption of normality.
- ACF plot has several correlation values outside of the confidence region, therefor the stochastic component of the series is not white noise. ACF plot reduces exponentially so time series may be stationary.
- Shapiro Wilks p-value of 0.6493, therefor fail to reject the null hypothesis that the stochastic component of this model is normally distributed.

## Conclusion

The superior model for this data is the quadratic model. The R-squared value and Shapiro-Wilks p-value for this model are higher than the linear model.