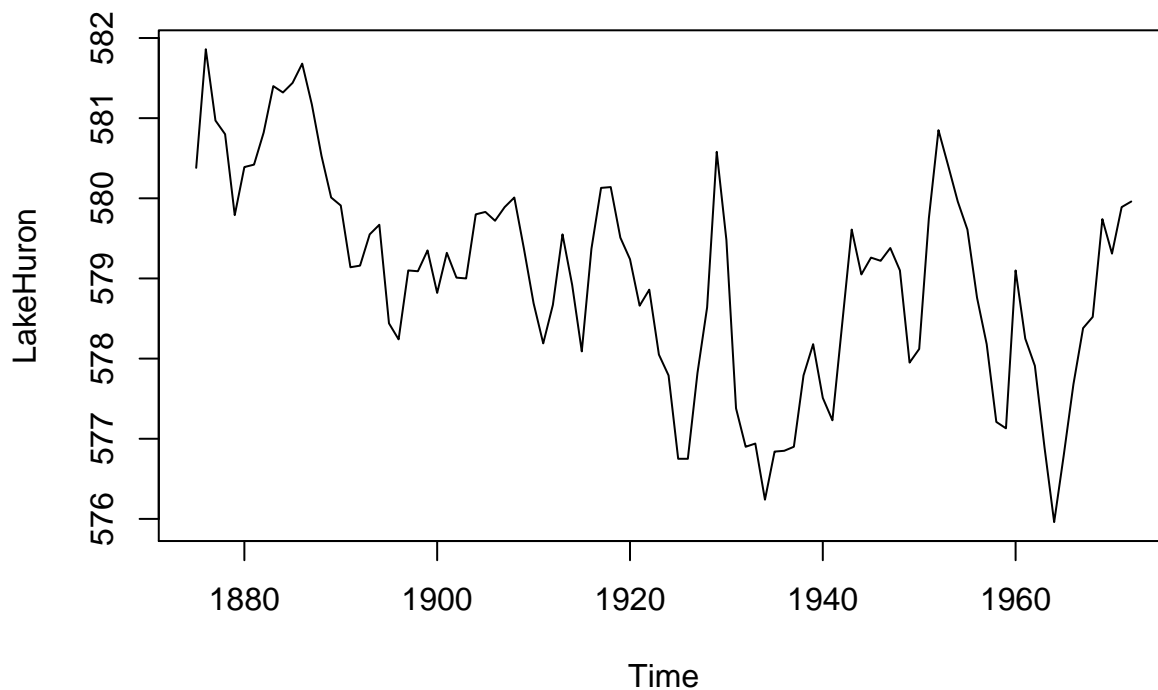


Homework 2

1.A

```
plot(LakeHuron)
```

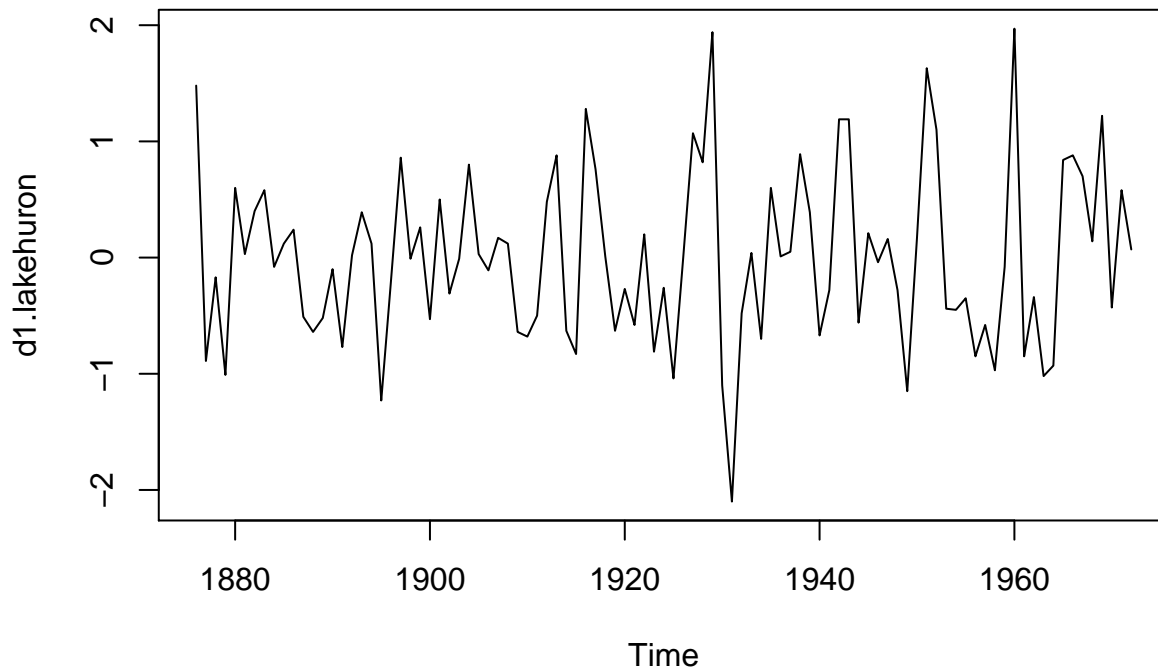


```
adf.test(LakeHuron)
```

```
##  
## Augmented Dickey-Fuller Test  
##  
## data: LakeHuron  
## Dickey-Fuller = -2.7796, Lag order = 4, p-value = 0.254  
## alternative hypothesis: stationary
```

The p-value from the Dickey-Fuller test confirms that we need to difference the time-series in order to make the data stationary.

```
d1.lakehuron <- diff(LakeHuron)  
plot(d1.lakehuron)
```



```
adf.test(d1.lakehuron)
```

```
## Warning in adf.test(d1.lakehuron): p-value smaller than printed p-value
```

```
##
```

```
## Augmented Dickey-Fuller Test
```

```
##
```

```
## data: d1.lakehuron
```

```
## Dickey-Fuller = -5.4687, Lag order = 4, p-value = 0.01
```

```
## alternative hypothesis: stationary
```

According the Dickey-Fuller test, the Lake Huron time series is stationary given that null hypothesis is rejected at the 5% significance level.

1.B

```
(m1 <- arima(d1.lakehuron, order = c(1,0,0), method = 'ML'))
```

```
##
```

```
## Call:
```

```
## arima(x = d1.lakehuron, order = c(1, 0, 0), method = "ML")
```

```
##
```

```
## Coefficients:
```

```
##          ar1  intercept
```

```
##          0.1362   -0.0018
```

```
## s.e.  0.1022    0.0867
```

```
##
```

```
## sigma^2 estimated as 0.5452: log likelihood = -108.23, aic = 222.45
```

1.C

```
(m2 <- arima(d1.lakehuron, order = c(2,0,0), method = 'ML'))

##
## Call:
## arima(x = d1.lakehuron, order = c(2, 0, 0), method = "ML")
##
## Coefficients:
##          ar1      ar2  intercept
##      0.1726 -0.2234   -0.0044
## s.e.  0.1012   0.1015    0.0698
##
## sigma^2 estimated as 0.5188:  log likelihood = -105.87,  aic = 219.74
```

1.D

```
D <- -2*(m1$loglik - m2$loglik)
pval <- 1-pchisq(D,1)
print(c("Test Statistic:",round(D,4),"P-value:",round(pval,4)))

## [1] "Test Statistic:" "4.7148"          "P-value:"          "0.0299"
```

1.E

The likelihood ratio test conducted in section D suggests that the AR(2) model doesn't perform significantly better than the AR(1). The variance and aic value for the AR(2) model however, show that it fits comfortably better than the AR(1). Since AR(2) is not terribly more complex than the AR(1) model, AR(2) would be a better option.

1.F

Mean of Zero

```
e <- m2$residuals
r <- e/sqrt(m2$sigma2)
t.test(r)

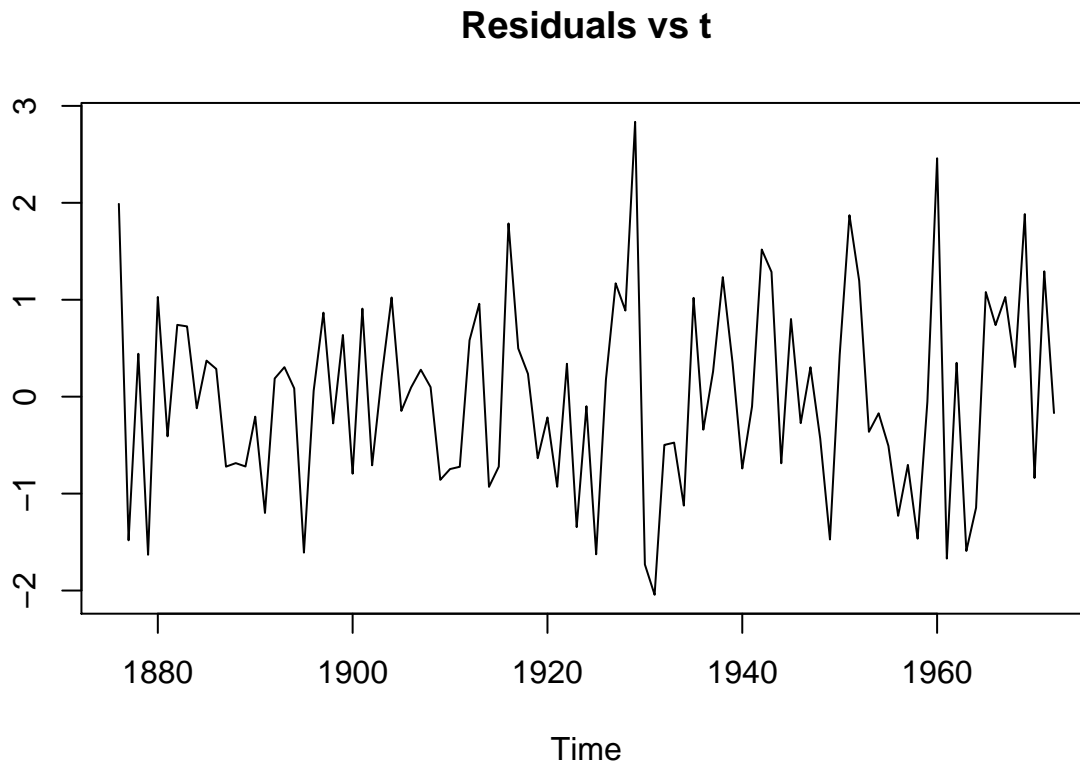
##
## One Sample t-test
##
## data:  r
## t = -0.013999, df = 96, p-value = 0.9889
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.2040202  0.2011626
## sample estimates:
## mean of x
```

```
## -0.0014288
```

Given that the p-value is near 1, the null hypothesis that the mean is equal to zero is not rejected. We can safely assume the mean of the residuals is zero.

Homoskedasticity

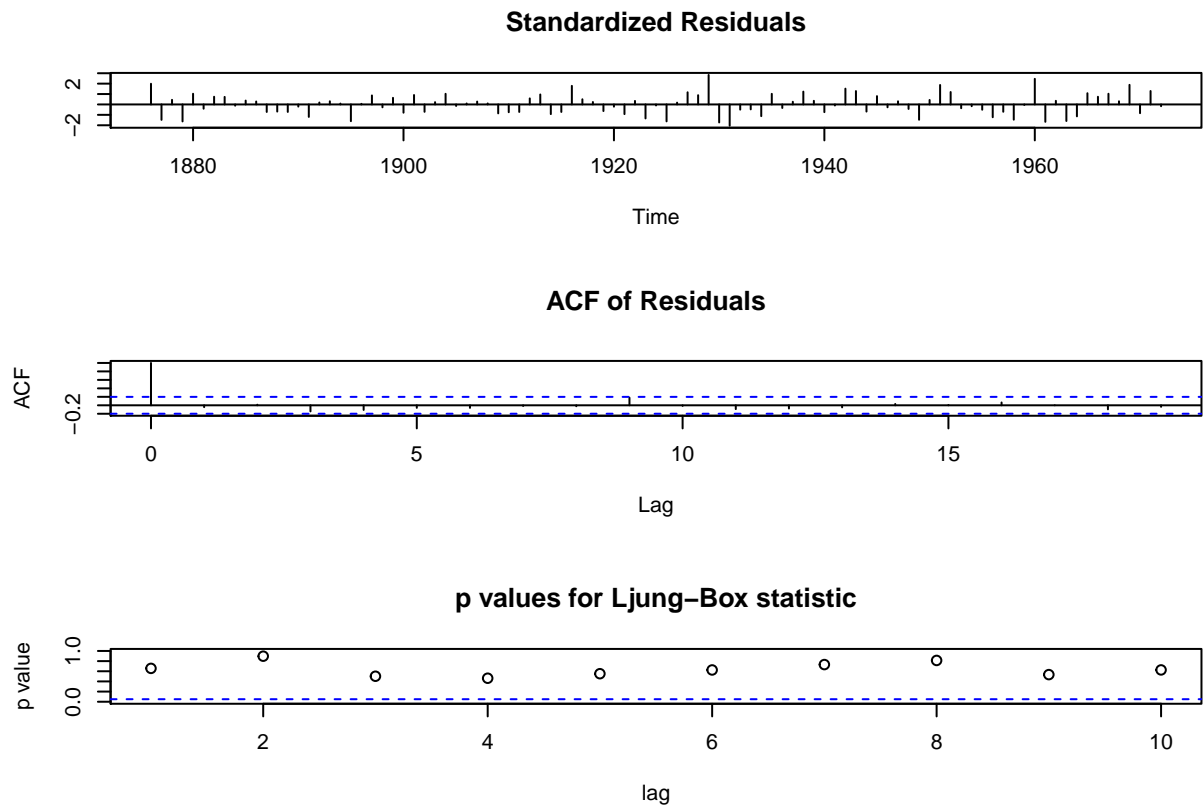
```
plot(r, main="Residuals vs t", ylab="")
```



From the residual plot above, heteroskedasticity doesn't seem to be much of a problem since there is no apparent increase in variation with shifts in t .

Auto-correlatedness

```
tsdiag(m2)
```

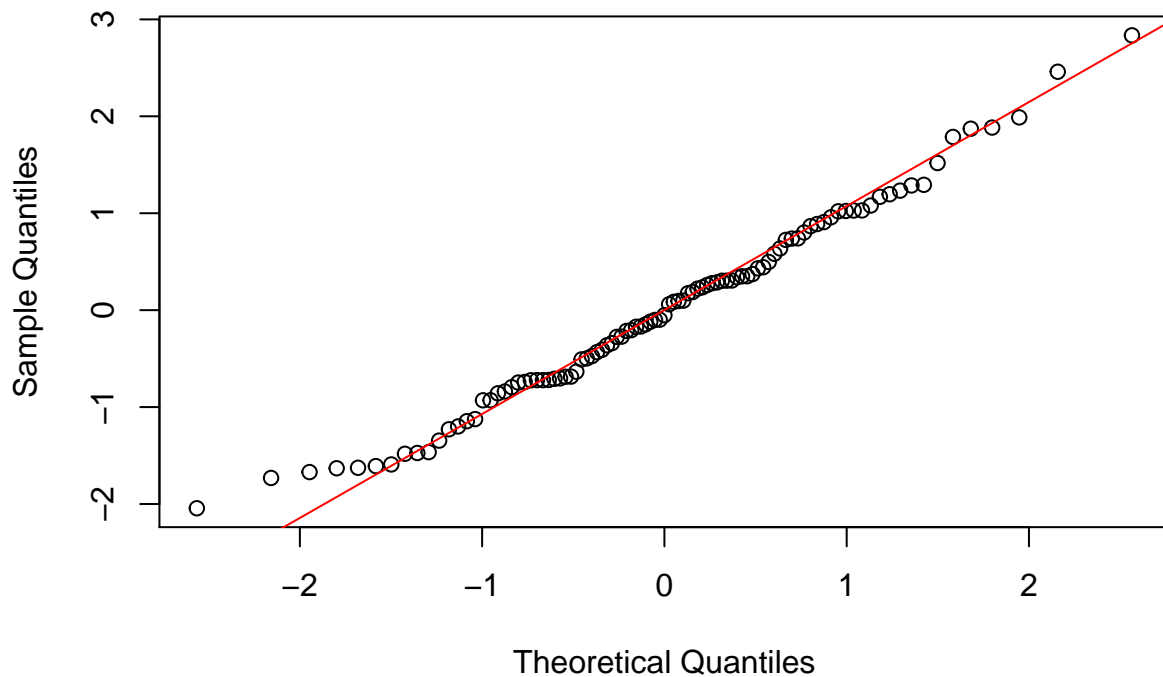


As shown in both the ACF plot and Ljung-Box statistic plot, the standardized residuals are not correlated at lags greater than 0. This makes us happy.

Normality

```
qqnorm(r, main="QQ-plot of Residuals")
qqline(r, col = "red")
```

QQ-plot of Residuals



For the most part, the residuals fall on the qqline but a shapiro wilk's test for normality will more definitively tell us if the residuals come from a normal distribution.

```
shapiro.test(r)
```

```
##  
##  Shapiro-Wilk normality test  
##  
## data:  r  
## W = 0.98644, p-value = 0.4233
```

Since the p-value is greater than any usual significance level, the null hypothesis that the data come from a normal distribution is not rejected, therefore we can safely assume that the data come from a normal distribution.

Overall Residual Assessment

The residuals do not violate any of the assumptions necessary to be comfortable with the fit of our model. This also gives us confidence that our model will do well during forecasting.

2