STAT4181 HW5 Min Yang

A. Seasonal ARIMA

1.

```r
library(zoo)
```

```
##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```r
library(rdatamarket)
library(forecast)
library(astsa)
```

```
##
## Attaching package: 'astsa'

## The following object is masked from 'package:forecast':
##
##     gas
```

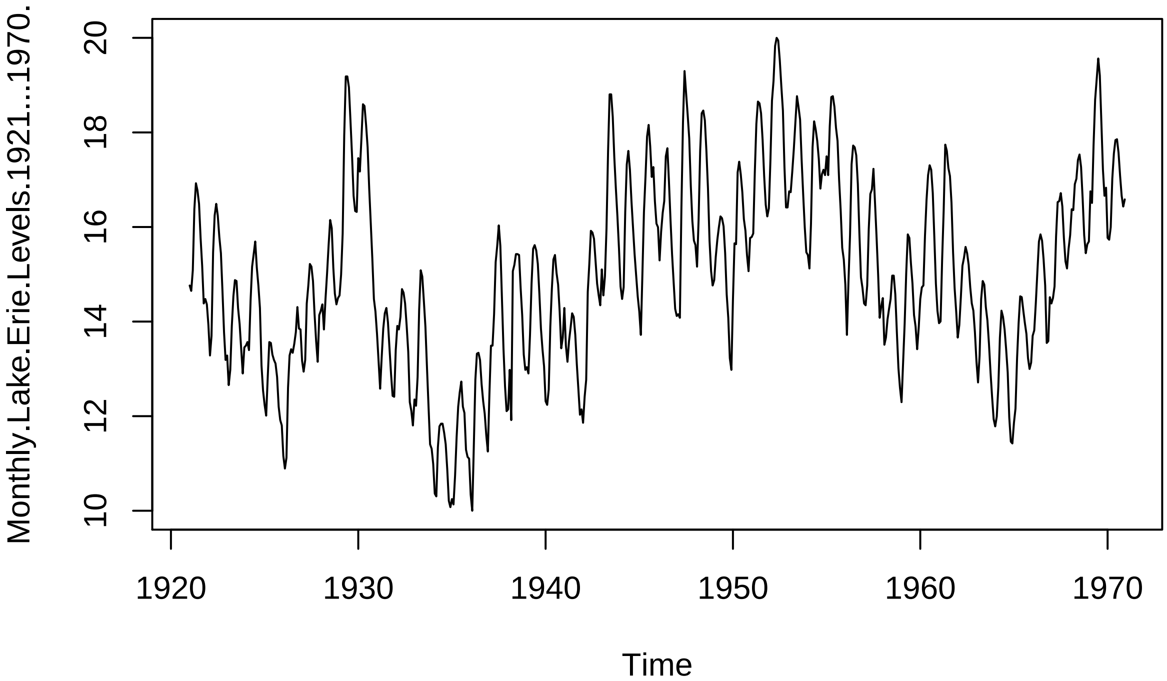
3.

```r
data<-dmseries("https://datamarket.com/data/set/22pw/monthly-lake-erie-levels-1921-1970#!ds=22pw&displa
str(data)
```

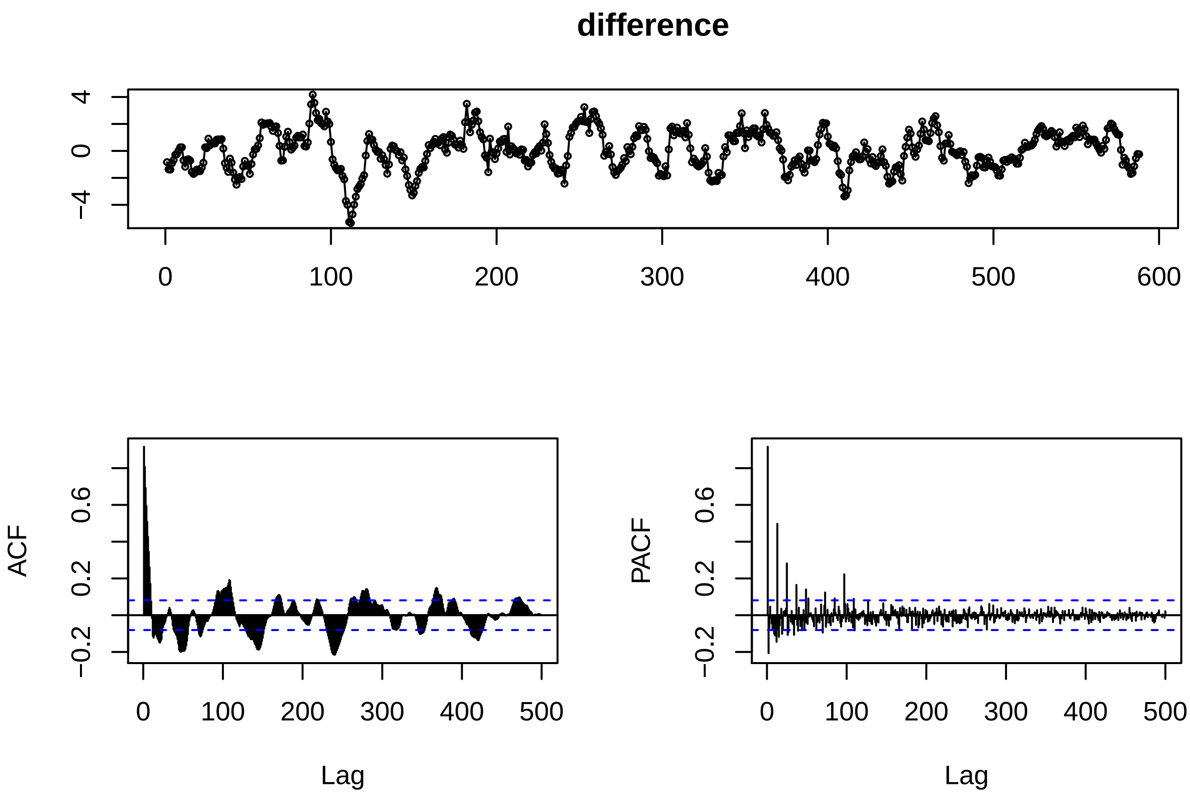
```
## 'zoo' series from Jan 1921 to Dec 1970
##   Data: num [1:600, 1] 14.8 14.6 15.1 16.4 16.9 ...
##  - attr(*, "dimnames")=List of 2
##   ..$ : NULL
##   ..$ : chr "Monthly.Lake.Erie.Levels.1921...1970."
##   Index: Class 'yearmon'  num [1:600] 1921 1921 1921 1921 1921 ...
```

4.

```r
plot.ts(data)
```




```
difference<-diff(data,12)
tsdisplay(difference,lag.max=500)
```



The acf plot suggests there that the data are highly correlated, and there are very strong autocorrelation. Both positive and negative autocorrelation occure, with a negative followed by a positive. A negative autocorrelation means the lake level decrease and positive autocorrelation means the lake level increase. So there is a seasonality.

5.

```r
auto.arima(data, stepwise = FALSE)
```

```
## Series: data
## ARIMA(0,1,2)(2,0,0)[12]
##
## Coefficients:
##          ma1     ma2    sar1    sar2
##       0.2599  0.1074  0.3076  0.3771
## s.e.  0.0432  0.0420  0.0383  0.0387
##
## sigma^2 estimated as 0.2105:  log likelihood=-384.81
## AIC=779.62   AICc=779.72   BIC=801.6
```

The best model is ARIMA(0,1,2)(2,0,0)[12].

6.

```r
sarima(data,p = 0,d = 1,q = 2,P=2,D = 0,Q = 0,S = 12,details=FALSE)
```

```
## $fit
##
## Call:
## stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
##     Q), period = S), xreg = constant, optim.control = list(trace = trc, REPORT = 1,
##     reltol = tol))
##
## Coefficients:
##          ma1     ma2    sar1    sar2  constant
##       0.2599  0.1074  0.3076  0.3771    0.0006
## s.e.  0.0432  0.0420  0.0383  0.0387    0.0760
##
## sigma^2 estimated as 0.2091:  log likelihood = -384.81,  aic = 781.62
##
## $degrees_of_freedom
## [1] 594
##
## $ttable
##          Estimate     SE t.value p.value
## ma1        0.2599 0.0432  6.0221  0.0000
## ma2        0.1074 0.0420  2.5577  0.0108
## sar1       0.3076 0.0383  8.0291  0.0000
## sar2       0.3771 0.0387  9.7355  0.0000
## constant   0.0006 0.0760  0.0080  0.9936
##
## $AIC
## [1] -0.5482466
##
## $AICc
## [1] -0.5446772
##
## $BIC
## [1] -1.511605
```

The estimated generating equation:

$$(1 - 0.3076B^{12} - 0.3771B^{24})(1 - 0.2599B - 0.1074B^2)\nabla^1 x_t = 0.0006 + w_t,$$

B. Linear Regression

1.

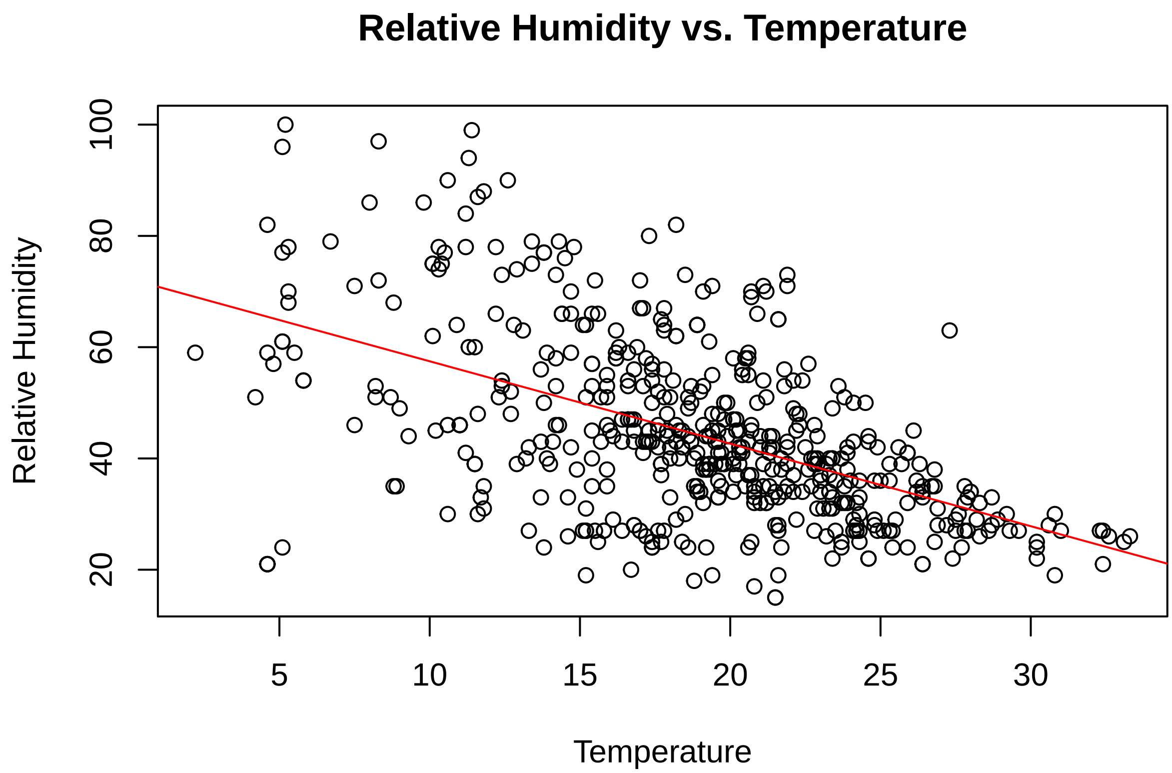
```
forest<-read.csv("~/desktop/forestfires.csv",header=TRUE)
```

2.

```
lm_model<-lm(RH~temp,forest)
```

3.

```
plot(forest$temp,forest$RH,xlab="Temperature", ylab="Relative Humidity",
     main="Relative Humidity vs. Temperature")
abline(lm_model,col=2)
```



4.

```
summary(lm_model)
```

```
##
## Call:
## lm(formula = RH ~ temp, data = forest)
##
## Residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -44.465  -8.083  -0.905   7.176  43.613
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  72.2828     2.0789   34.77   <2e-16 ***
## temp         -1.4820     0.1052  -14.09   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.88 on 515 degrees of freedom
## Multiple R-squared:  0.2781, Adjusted R-squared:  0.2767
## F-statistic: 198.4 on 1 and 515 DF,  p-value: < 2.2e-16
```

The estimated equation is RH=72.2828-1.482temp. With p-value <0.0001, the regression model is significant.But the r-square value shows the model only explains 27.81% of the data. Since this is a simple linear regression model, there should be other varaibles that affect the relative humidity.