lab4

2023-10-27

1. Reading the Data

We introduce the Time Series Data Library (TSDL) created by Professor Rob Hyndman.(https://rdrr.io/github/FinYang/tsdl/f/vignettes/tsdl.Rmd) Run the following code to import quarterly Iowa nonfarm income in R:

```r
#uncomment the next three lines when you first install tsdl library
#install.packages("devtools")
#install.packages("forecast")
#devtools::install_github("FinYang/tsdl")
library(tsdl)
library(forecast)
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```r
meta_tsdl$description[[1]]
```

```
## [1] "Quarterly Iowa nonfarm income (1948 - 1979)"
```

```r
iowa.ts <- tsdl[[1]]
head(iowa.ts)
```

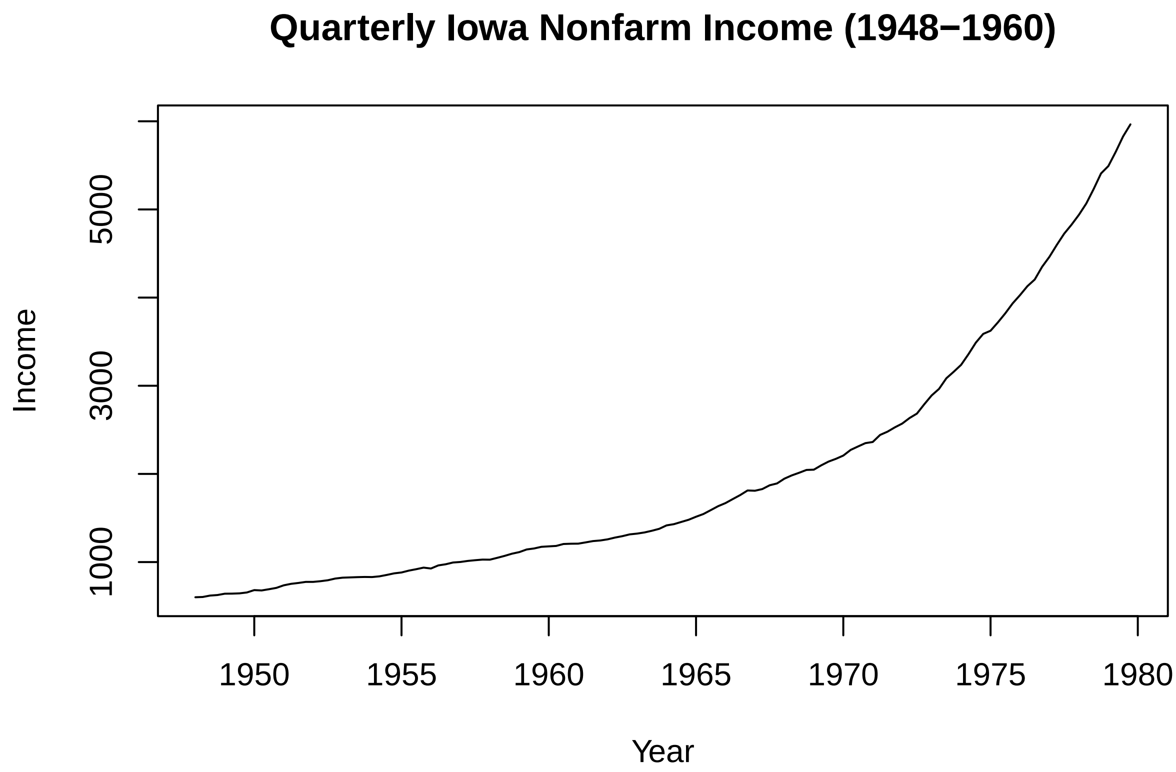
```
##      Qtr1 Qtr2 Qtr3 Qtr4
## 1948  601  604  620  626
## 1949  641  642
```

2. Data Analysis

a. Plot the time series. What do you notice? Does the variance change over time? Is their a trend and/or seasonal components?

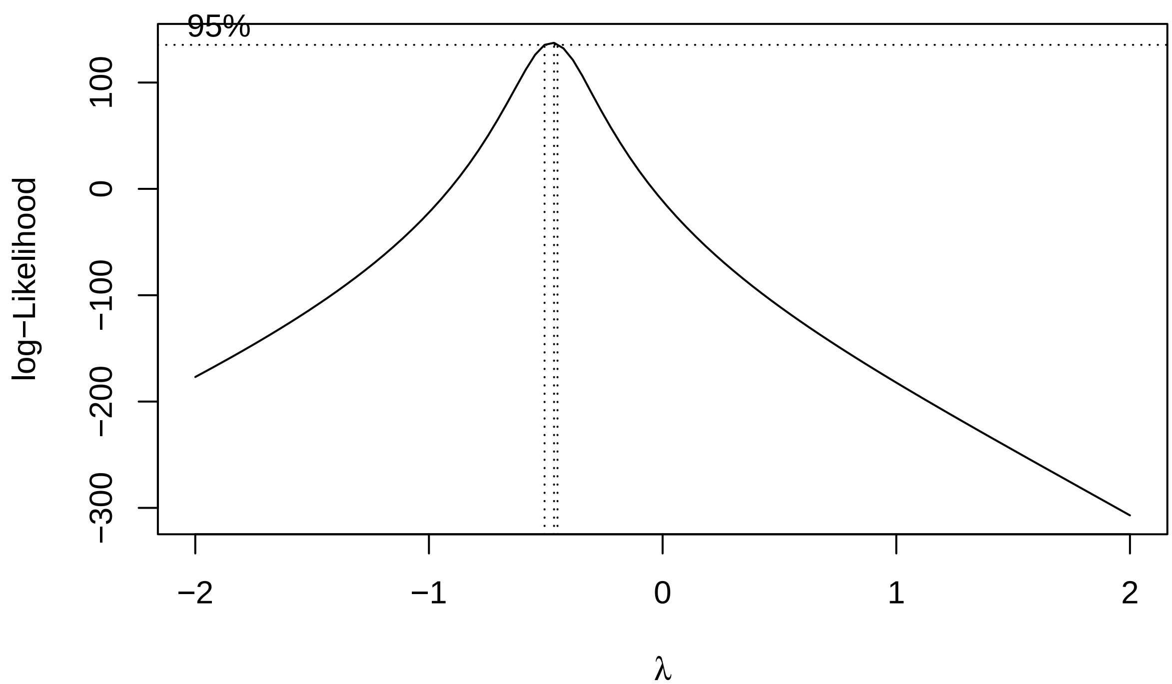
- There is an increasing trend in the data.
- The variability of the data changing over time
- There is also a strong seasonal component.

```r
plot(iowa.ts, xlab="Year", ylab="Income", main="Quarterly Iowa Nonfarm Income (1948-1960)")
```



b. Apply a Box-Cox transformation to the time series, find the optimal lambda, transform the data, and re-plot the transformed time series vs. original time series.

```
library(MASS)
t = 1:length(iowa.ts)
fit = lm(iowa.ts ~ t)
bcTransform = boxcox(iowa.ts ~ t,plotit = TRUE)
```



The dashed vertical lines in the plot above (which is created automatically using the argument plotit = TRUE) correspond to a 95% confidence interval for the true value of lambda in the Box-Cox transformation. If the confidence interval includes lambda = 0, then the we can try log transformation (special case of Box-Cox transformation when lambda = 0, given by Y_t = log(X_t); otherwise, the Box-Cox transformation for stabilizing the variance is given by:

Y_t = 1/lambda X_t^{lambda} as we implement in the code below:

```
lambda = bcTransform$x[which(bcTransform$y == max(bcTransform$y))]
iowa.bc = (1/lambda)*(iowa.ts^lambda - 1)
```

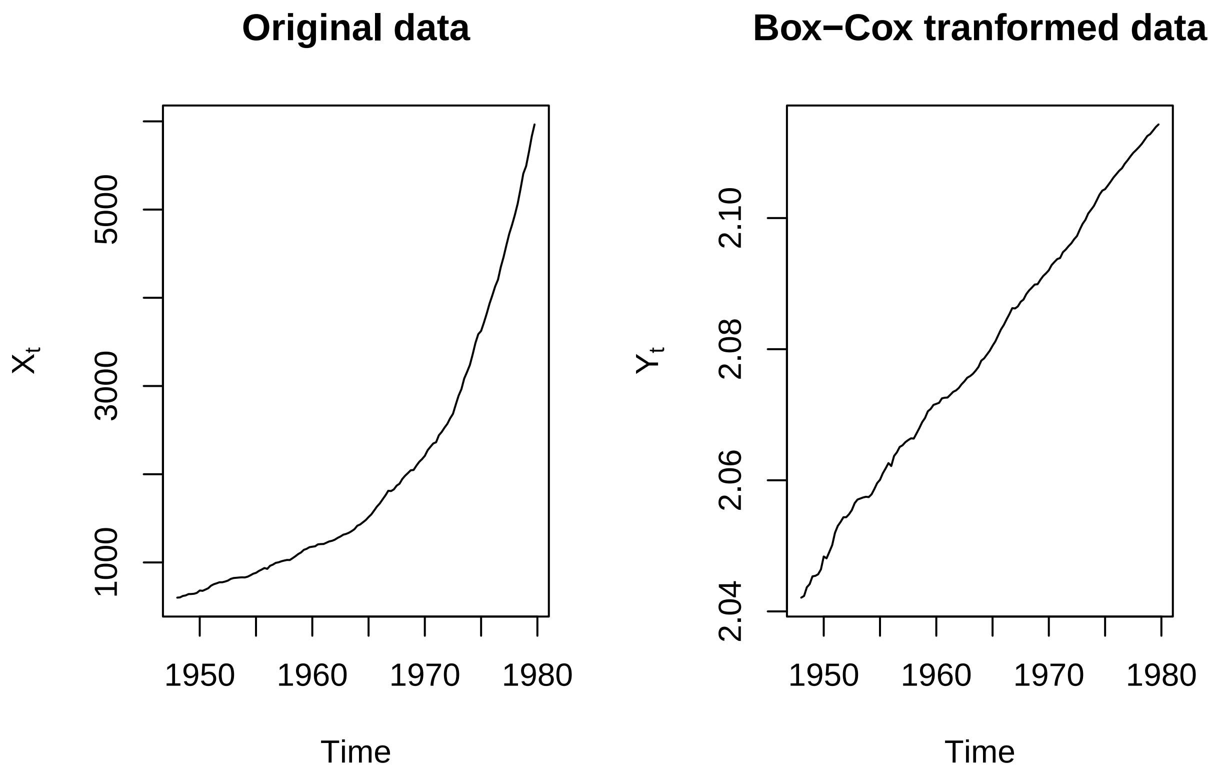
To print out the optimal lambda:

```
lambda
```

```
## [1] -0.4646465
```

We now plot the original data vs Box-Cox transformed data:

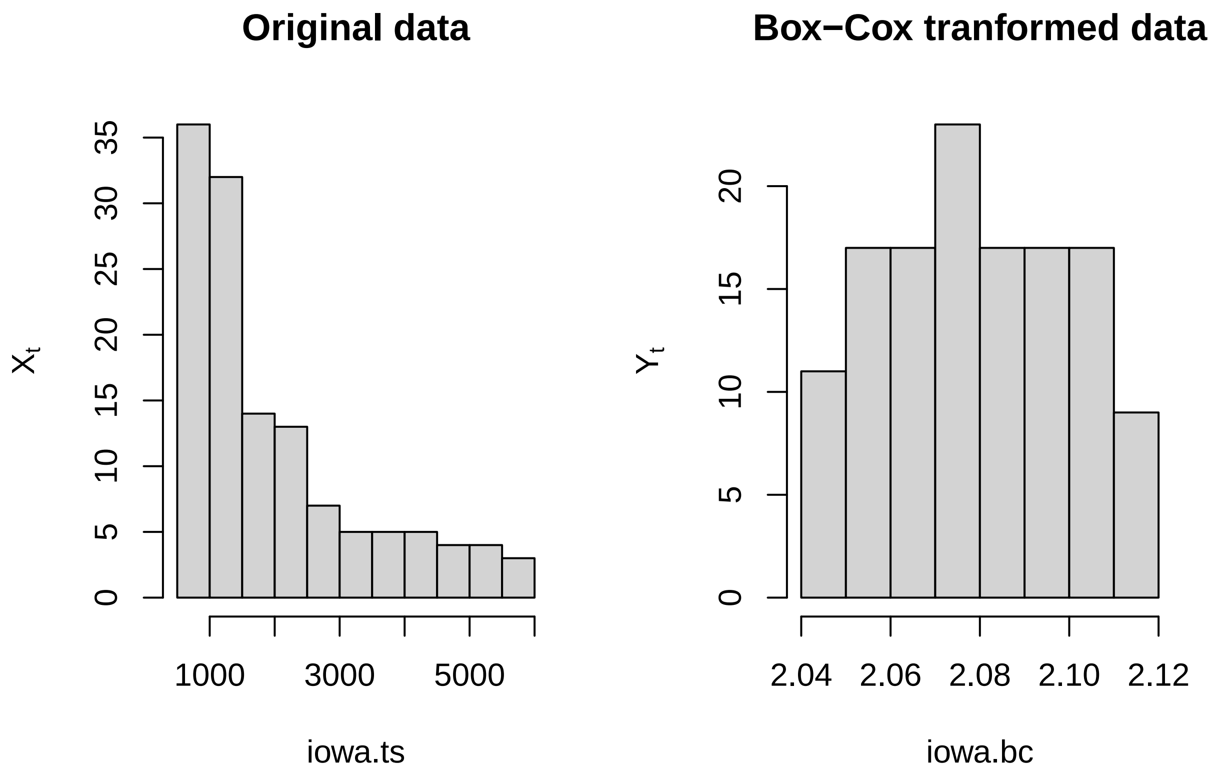
```
op <- par(mfrow = c(1,2))
ts.plot(iowa.ts,main = "Original data",ylab = expression(X[t]))
ts.plot(iowa.bc,main = "Box-Cox tranformed data", ylab = expression(Y[t]))
```




```r
par(op)
```

c. Calculate the sample variance and examine the sample ACF of the transformed data (set max.lag = 48 or 60). What do you notice? Access if there is trend in the data. If yes, remove trend components by using the diff() at lag=1. Plot the differenced time series. Does it look stationary? Re-calculate the sample variance. From the two plots above, the transformed data has a more stable variance acorss time

```r
op <- par(mfrow = c(1,2))
hist(iowa.ts,main = "Original data",ylab = expression(X[t]))
hist(iowa.bc,main = "Box-Cox tranformed data", ylab = expression(Y[t]))
```




```r
par(op)
```

```r
# Calculate the sample variance and plot the acf/pacf
var(iowa.ts)
```

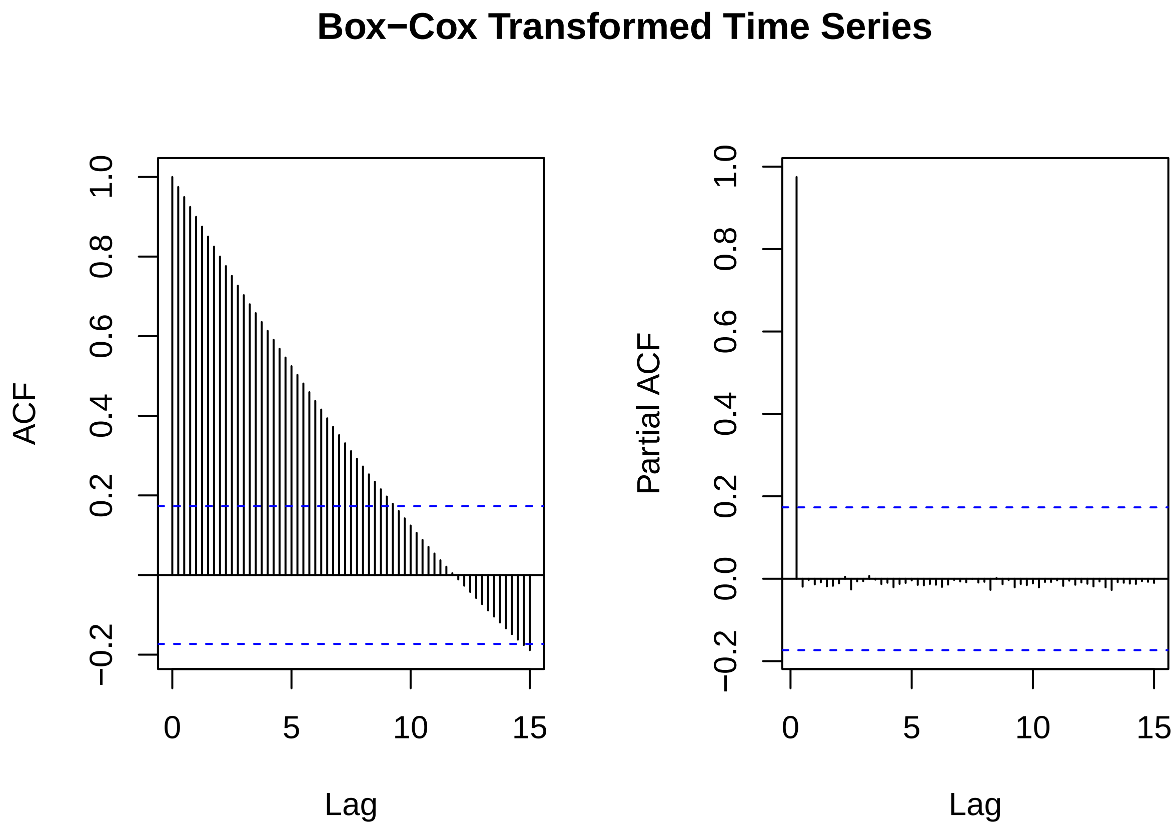
```
## [1] 1940784
```

```r
var(iowa.bc)
```

```
## [1] 0.0004190696
```

- From the ACF graph, it shows a slow, graduate decay which also indicate a trend or non-stationary of data.

```r
op = par(mfrow = c(1,2))
acf(iowa.bc,lag.max = 60,main = "")
pacf(iowa.bc,lag.max = 60,main = "")
title("Box-Cox Transformed Time Series", line = -1, outer=TRUE)
```



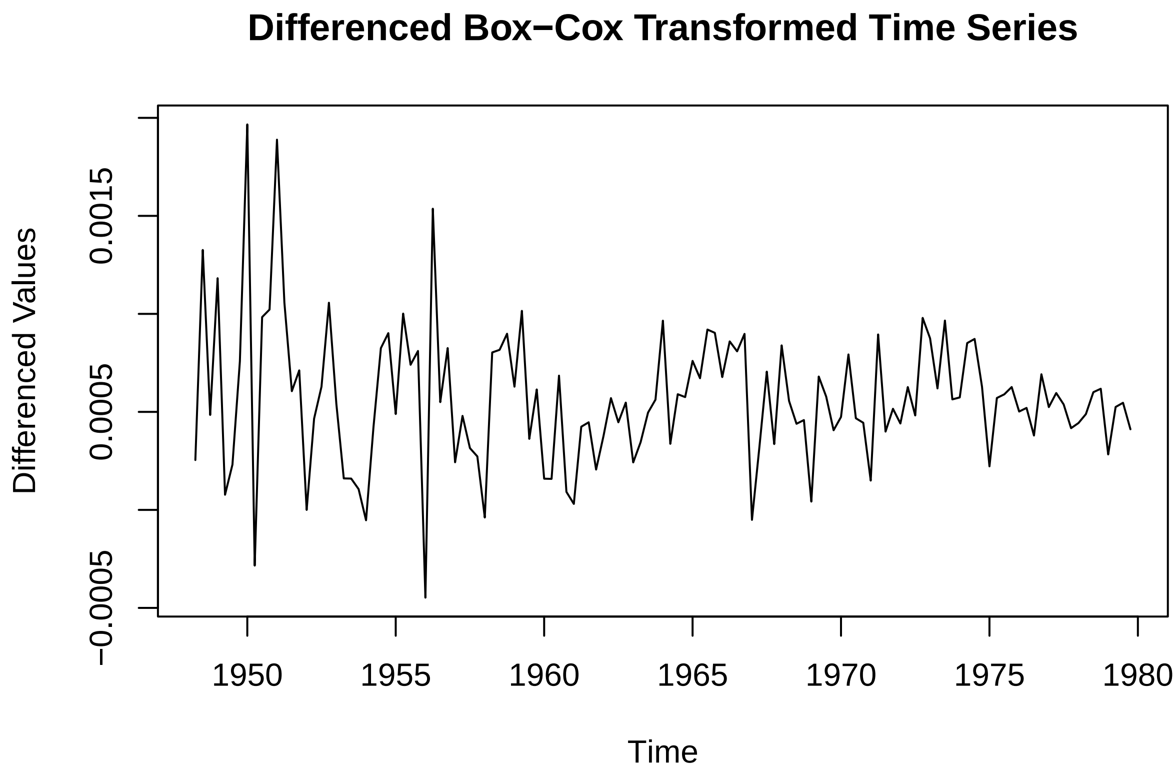

```
par(op)
```

To remove the trend using diff.

```
iowa.bc.diff = diff(iowa.bc, lag = 1)
```

Plot the differenced time series.

```
plot(iowa.bc.diff, main = "Differenced Box-Cox Transformed Time Series", ylab = "Differenced Values")
```



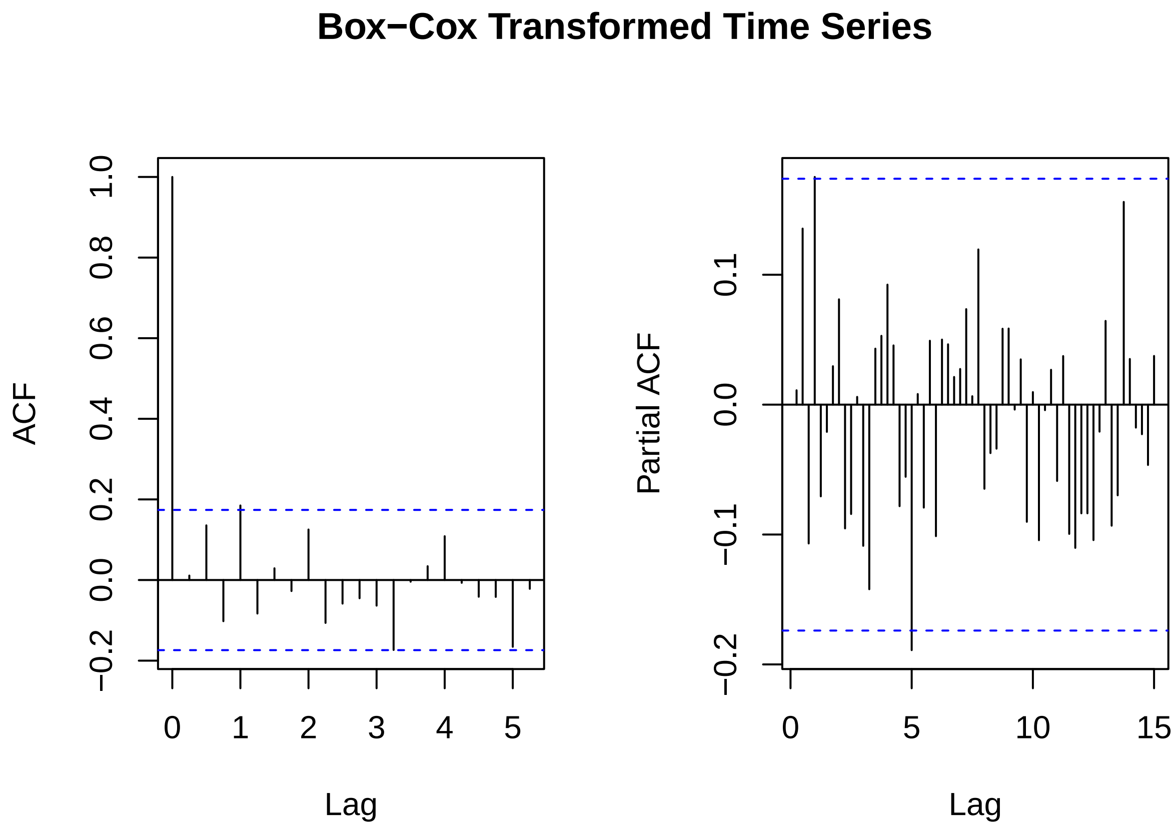
A stationary series should not display any visible trend and should have a constant variance over time. After we remove the trend, the iowa looks more stationary, and its variance is really close to zero.

```
var(iowa.bc.diff)
```

```
## [1] 1.266047e-07
```

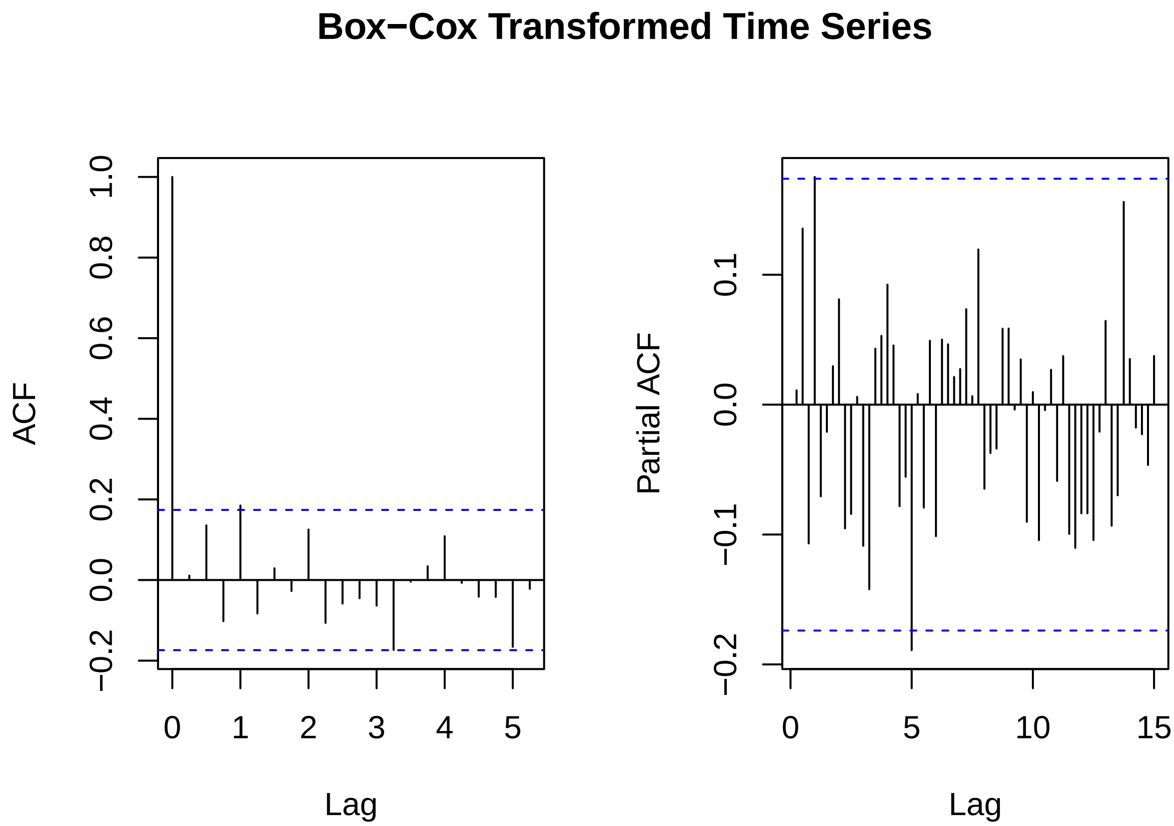
d. If there is seasonality, determine the seasonal period and remove seasonal components by differencing the time series data from (c) using the diff() function. Re-calculate the sample variance. After plotting acf for new time series iowa.bc.diff, it looks like there is no seasonality

```
op = par(mfrow = c(1,2))
acf(iowa.bc.diff,max = 60,main = "")
pacf(iowa.bc.diff,lag.max = 60,main = "")
title("Box-Cox Transformed Time Series", line = -1, outer=TRUE)
```



e. For the stationary data you get, examine the sample ACF and sample PACF and pick significant lags for sample ACF and sample PACF.

```
op = par(mfrow = c(1,2))
acf(iowa.bc.diff,max = 60,main = "")
pacf(iowa.bc.diff,lag.max = 60,main = "")
title("Box-Cox Transformed Time Series", line = -1, outer=TRUE)
```



The significant lag in ACF is only lag 0, and there is no significant lag in PACF. In this case, we can determine that there is no significant autocorrelation in the stationary data like that.