ENV 790.30 - Time Series Analysis for Energy Data | Spring 2025

Assignment 3 - Due date 02/04/25

Jingze Dai

Directions

You should open the .rmd file corresponding to this assignment on RStudio. The file is available on our class repository on Github.

Once you have the file open on your local machine the first thing you will do is rename the file such that it includes your first and last name (e.g., "LuanaLima_TSA_A03_Sp25.Rmd"). Then change "Student Name" on line 4 with your name.

Then you will start working through the assignment by **creating code and output** that answer each question. Be sure to use this assignment document. Your report should contain the answer to each question and any plots/tables you obtained (when applicable).

Please keep this R code chunk options for the report. It is easier for us to grade when we can see code and output together. And the tidy.opts will make sure that line breaks on your code chunks are automatically added for better visualization.

When you have completed the assignment, **Knit** the text and code into a single PDF file. Submit this pdf using Sakai.

Questions

Consider the same data you used for A2 from the spreadsheet "Table_10.1_Renewable_Energy_Production_and_Consumpti The data comes from the US Energy Information and Administration and corresponds to the December 2024 **Monthly** Energy Review. Once again you will work only with the following columns: Total Renewable Energy Production and Hydroelectric Power Consumption. Create a data frame structure with these two time series only.

R packages needed for this assignment:"forecast","tseries", and "Kendall". Install these packages, if you haven't done yet. Do not forget to load them before running your script, since they are NOT default packages.\

```
#Load/install required package here
list_of_lib <- c("forecast","tseries","Kendall","dplyr","openxlsx","ggplot2")
for (i in list_of_lib){
  library(i,  character.only = TRUE)
}
```

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

##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

Importing data

```r
# importing dataset
energy_data <- read.xlsx(
  xlsxFile="../Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source.xlsx",
  sheet = "Monthly Data", startRow = 13, colNames = FALSE)

read_col_names  <- read.xlsx(
  xlsxFile="../Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source.xlsx",
  sheet = "Monthly Data", rows = 11, colNames = FALSE)

colnames(energy_data) <- read_col_names
energy_data$Data <- convertToDate(energy_data$Month)

# selecting columns
power_data <- energy_data %>%
  select("Total Renewable Energy Production",
         "Hydroelectric Power Consumption")

# transforming to time series
ts_power_data <- ts(power_data, start=c(1973,1), frequency=12)
```

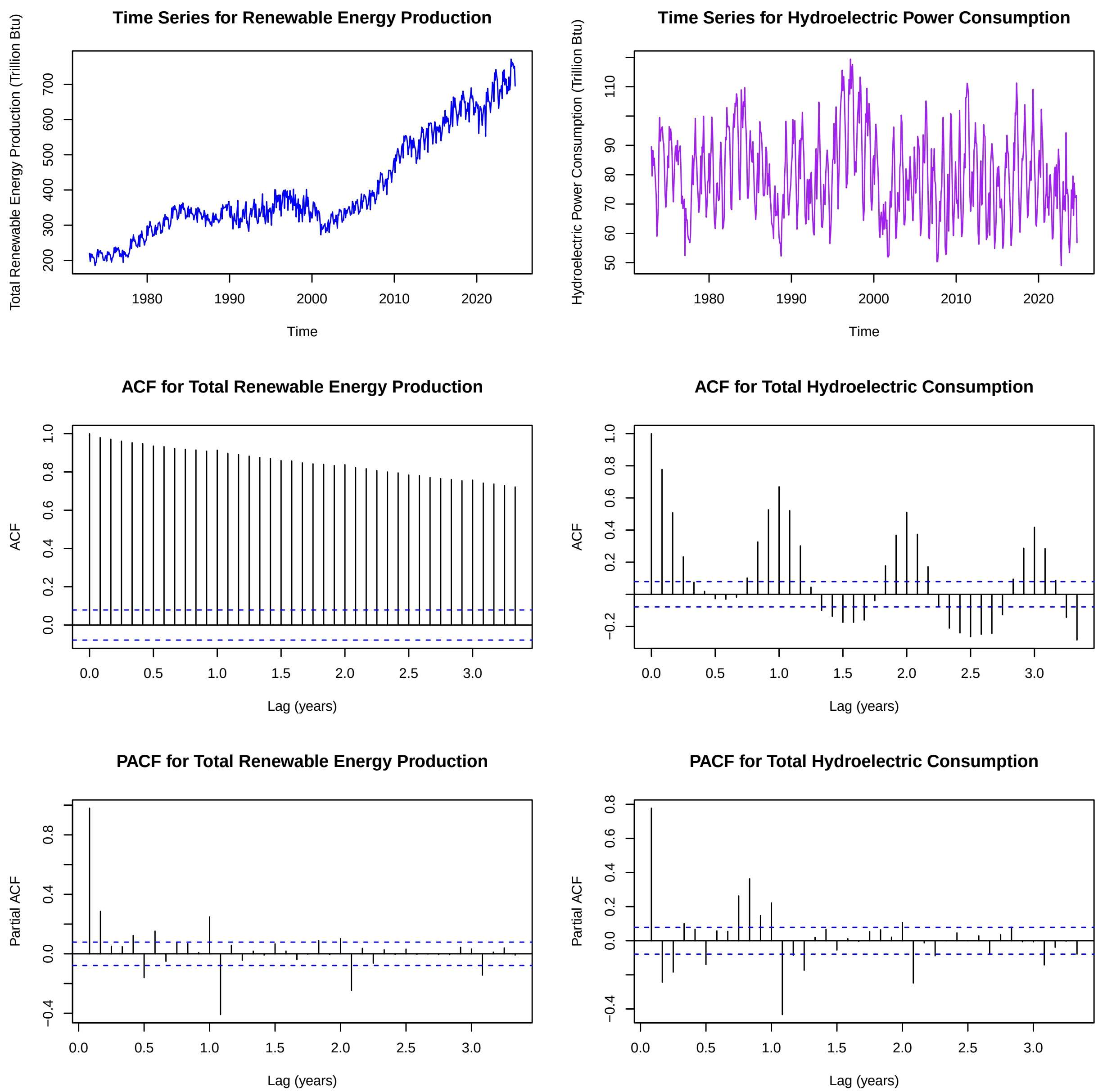
##Trend Component

Q1

For each time series, i.e., Renewable Energy Production and Hydroelectric Consumption create three plots: one with time series, one with the ACF and with the PACF. You may use the some code form A2, but I want all the three plots side by side as in a grid. (Hint: use function `plot_grid()` from the `cowplot` package)

```r
par(mfrow = c(3,2))

# time series plot
plot(ts_power_data[,1],
     type="l",col="blue",
     ylab="Total Renewable Energy Production (Trillion Btu)",
     main="Time Series for Renewable Energy Production")

plot(ts_power_data[,2],
     type="l",col="purple",
     ylab="Hydroelectric Power Consumption (Trillion Btu)",
     main="Time Series for Hydroelectric Power Consumption")

# ACF plot
```

```r
renewable_acf<-acf(ts_power_data[,1],lag.max=40,
                  main = "ACF for Total Renewable Energy Production",
                  type="correlation", plot=TRUE,
                  xlab="Lag (years)")

hydro_acf<-acf(ts_power_data[,2],lag.max=40,
              main = "ACF for Total Hydroelectric Consumption",
              type="correlation", plot=TRUE,
              xlab="Lag (years)")

# PACF plot
renewable_pacf<-pacf(ts_power_data[,1],lag.max=40,
                  main = "PACF for Total Renewable Energy Production",
                  plot=TRUE,xlab="Lag (years)")

hydro_pacf<-pacf(ts_power_data[,2],lag.max=40,
              main = "PACF for Total Hydroelectric Consumption",
              plot=TRUE,xlab="Lag (years)")
```



Q2

From the plot in Q1, do the series Total Renewable Energy Production and Hydroelectric Power Consumption appear to have a trend? If yes, what kind of trend?

Ans: The time series plot of the *total renewable energy production* shows a increasing trend that is relatively linear, while the plot of *total hydroelectric consumption* does not appear to have a significant trend. Even if there is a trend for the total hydroelectric consumption time series, it is likely to be non-linear.

Q3

Use the *lm()* function to fit a linear trend to the two time series. Ask R to print the summary of the regression. Interpret the regression output, i.e., slope and intercept. Save the regression coefficients for further analysis.

```r
nobs <- nrow(power_data)
t <- c(1:nobs)

renewable_linear_trend <- lm(power_data[,1] ~ t)
hydroelectric_linear_trend <- lm(power_data[,2] ~ t)

summary(renewable_linear_trend)
```

```
##
## Call:
## lm(formula = power_data[, 1] ~ t)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -151.11  -37.84   13.53   41.76  149.42
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 176.87293    4.96189   35.65   <2e-16 ***
## t             0.72393    0.01382   52.37   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 61.75 on 619 degrees of freedom
## Multiple R-squared:  0.8159, Adjusted R-squared:  0.8156
## F-statistic:  2743 on 1 and 619 DF,  p-value: < 2.2e-16
```

```r
summary(hydroelectric_linear_trend)
```

```
##
## Call:
## lm(formula = power_data[, 2] ~ t)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -29.995 -10.422  -0.720   9.161  39.624
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 82.96766    1.12339  73.855  < 2e-16 ***
## t           -0.01098    0.00313  -3.508 0.000485 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.98 on 619 degrees of freedom
## Multiple R-squared:  0.01949,    Adjusted R-squared:  0.01791
## F-statistic:  12.3 on 1 and 619 DF,  p-value: 0.0004848
```

```r
# saving the gradients and intercepts
renewable_intercept <- as.numeric(renewable_linear_trend$coefficients[1])
```

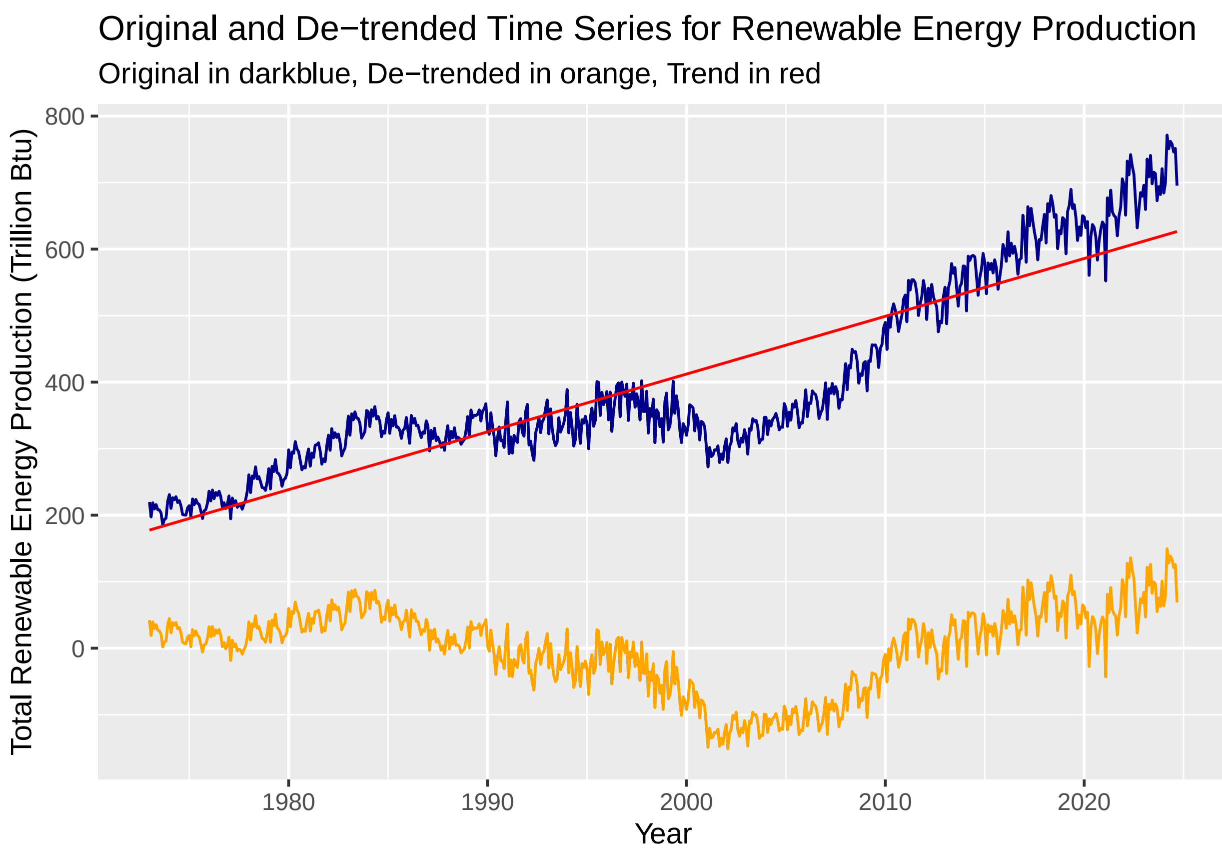
```r
renewable_gradient <- as.numeric(renewable_linear_trend$coefficients[2])
hydro_intercept <- as.numeric(hydroelectric_linear_trend$coefficients[1])
hydro_gradient <- as.numeric(hydroelectric_linear_trend$coefficients[2])
```

Ans: For renewable energy production: Intercept = 176.87293, Gradient = 0.72393, adjusted R-squared value = 0.8156. Both the intercept and gradient have significant p-values, and the positive gradient and the close-to-one adjusted R-squared value indicates a strong increasing linear trend. For hydroelectric consumption, Intercept = 82.96766, Gradient = -0.01098, adjusted R-squared value = 0.01791. Although the p-values for both the gradient and intercept are significant, the close-to-zero adjusted R-squared value and the negative gradient indicates a extremely weak decreasing linear trend.

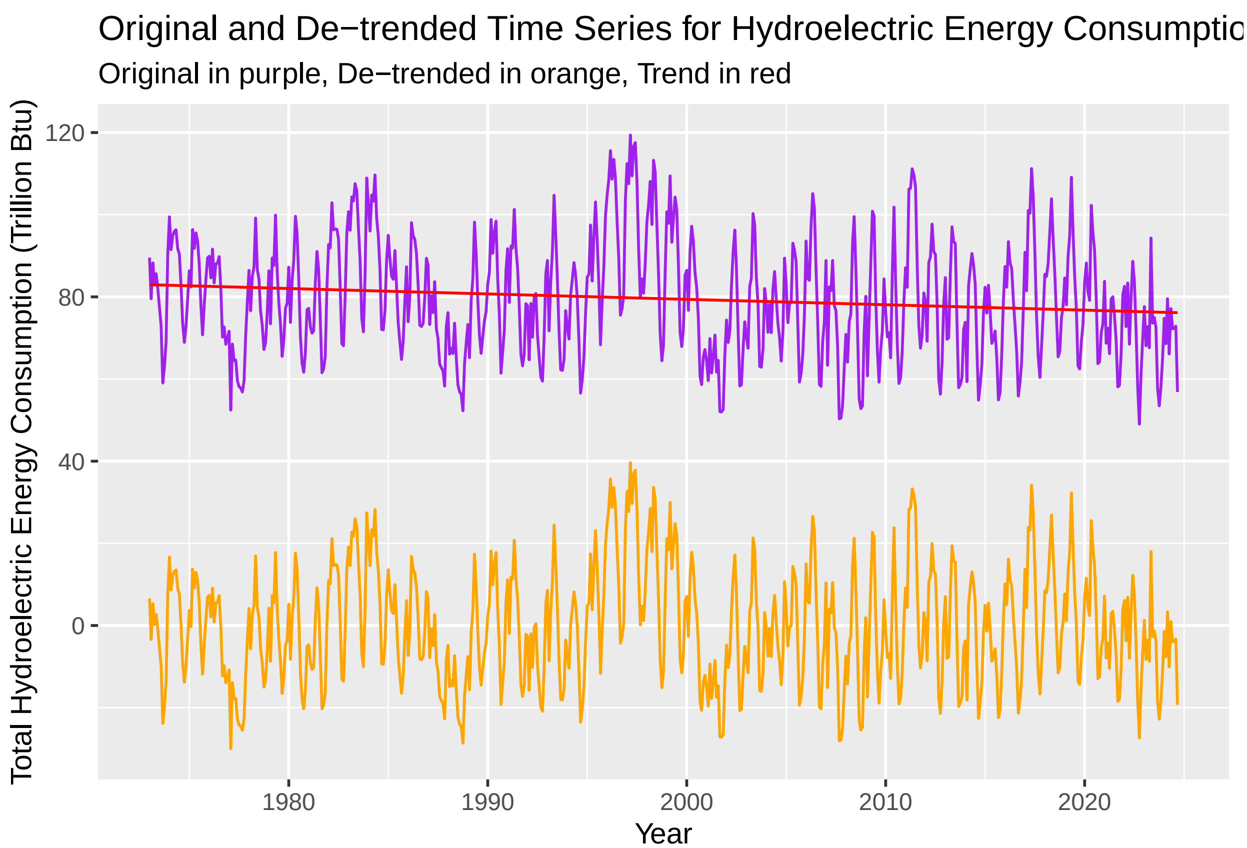
Q4

Use the regression coefficients from Q3 to detrend the series. Plot the detrended series and compare with the plots from Q1. What happened? Did anything change?

```r
# renewable
renewable_linear_trend <- renewable_intercept  + renewable_gradient * t
ts_renewable_linear <- ts(renewable_linear_trend,start=c(1973,1), frequency=12)

detrend_renewable <- power_data[,1] - renewable_linear_trend
ts_detrend_renewable <- ts(detrend_renewable, start = c(1973,1),frequency=12)

autoplot(ts_power_data[,1],color="darkblue")+
  autolayer(ts_detrend_renewable,series="Detrended",color="orange")+
  autolayer(ts_renewable_linear,series="Linear Component",color="red")+
  ggtitle(
    "Original and De-trended Time Series for Renewable Energy Production",
    subtitle = "Original in darkblue, De-trended in orange, Trend in red") +
  ylab("Total Renewable Energy Production (Trillion Btu)") + xlab("Year")
```




```r
# hydro
hydro_linear_trend <- hydro_intercept  + hydro_gradient * t
ts_hydro_linear <- ts(hydro_linear_trend,start=c(1973,1), frequency=12)

detrend_hydro <- power_data[,2] - hydro_linear_trend
ts_detrend_hydro <- ts(detrend_hydro, start = c(1973,1),frequency=12)

autoplot(ts_power_data[,2],color="purple")+
  autolayer(ts_detrend_hydro,series="Detrended",color="orange")+
  autolayer(ts_hydro_linear,series="Linear Component",color="red")+
  ggtitle(
    "Original and De-trended Time Series for Hydroelectric Energy Consumption",
    subtitle = "Original in purple, De-trended in orange, Trend in red") +
  ylab("Total Hydroelectric Energy Consumption (Trillion Btu)") + xlab("Year")
```



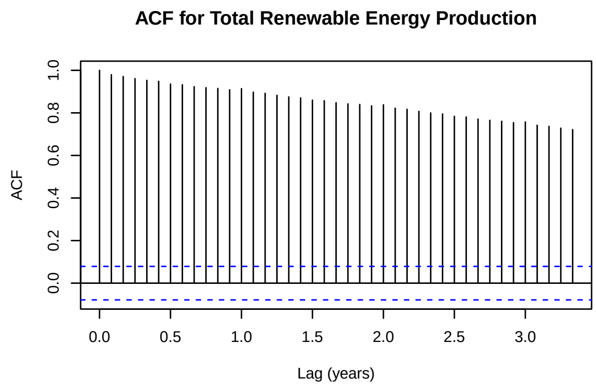
Ans: For the renewable plot, the de-trended time series changed from the original one, as the increasing trend disappeared. For the hydroelectric plot, the two time series look largely the same.

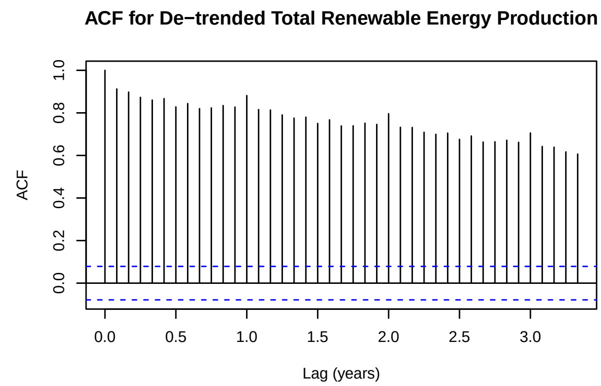
Q5

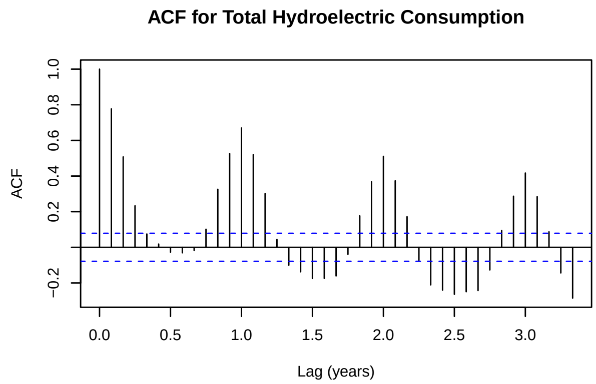
Plot ACF and PACF for the detrended series and compare with the plots from Q1. You may use plot_grid() again to get them side by side, but not mandatory. Did the plots change? How?

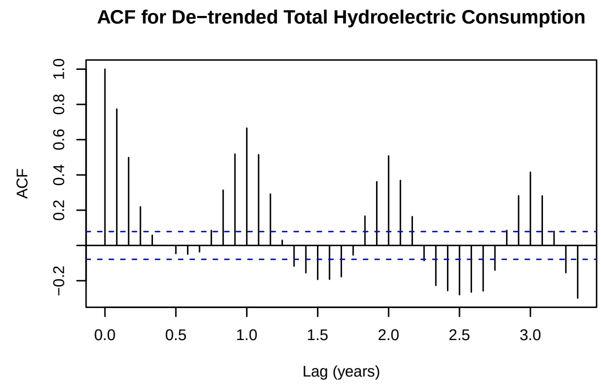
```r
par(mfrow = c(4,2))

# ACF plot
renewable_acf<-acf(ts_power_data[,1],lag.max=40,
                  main = "ACF for Total Renewable Energy Production",
                  type="correlation", plot=TRUE,
                  xlab="Lag (years)")

renewable_detrend_acf<-acf(ts_detrend_renewable,lag.max=40,
                  main = "ACF for De-trended Total Renewable Energy Production",
                  type="correlation", plot=TRUE,
                  xlab="Lag (years)")

hydro_acf<-acf(ts_power_data[,2],lag.max=40,
              main = "ACF for Total Hydroelectric Consumption",
              type="correlation", plot=TRUE,
              xlab="Lag (years)")
```

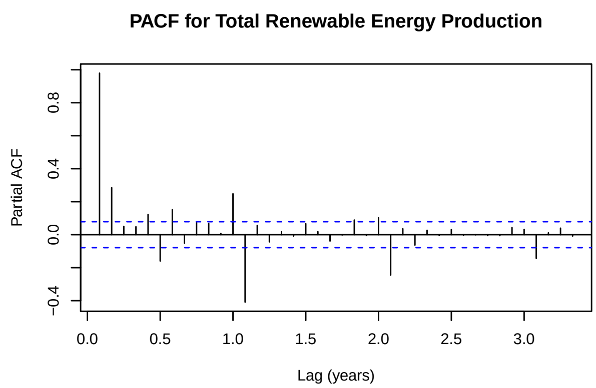
```r
hydro_detrended_acf<-acf(ts_detrend_hydro,lag.max=40,
              main = "ACF for De-trended Total Hydroelectric Consumption",
              type="correlation", plot=TRUE,
              xlab="Lag (years)")

# PACF plot
renewable_pacf<-pacf(ts_power_data[,1],lag.max=40,
                  main = "PACF for Total Renewable Energy Production",
                  plot=TRUE,xlab="Lag (years)")

renewable_detrended_pacf<-pacf(ts_detrend_renewable,lag.max=40,
                  main = "PACF for De-trended Total Renewable Energy Production",
                  plot=TRUE,xlab="Lag (years)")

hydro_pacf<-pacf(ts_power_data[,2],lag.max=40,
              main = "PACF for Total Hydroelectric Consumption",
              plot=TRUE,xlab="Lag (years)")

hydro_detrended_pacf<-pacf(ts_detrend_hydro,lag.max=40,
              main = "PACF for Total De-trended Hydroelectric Consumption",
              plot=TRUE,xlab="Lag (years)")
```

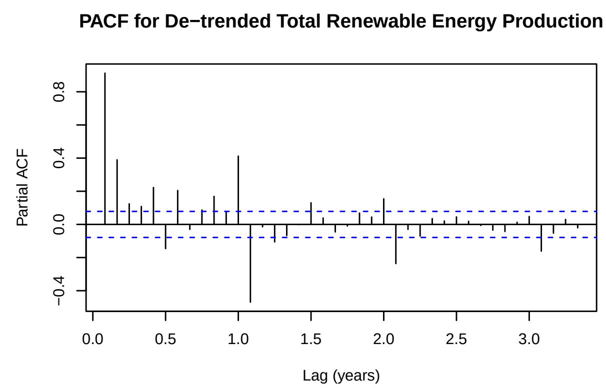


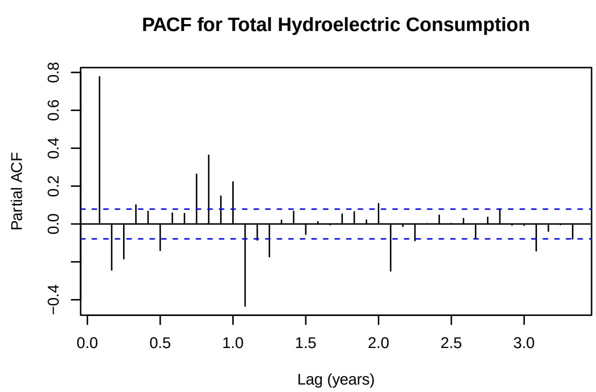


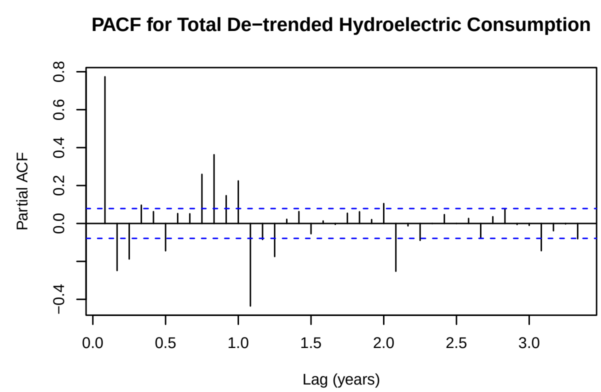












Ans: The ACF for renewable plot changed, and the ACF values decreased by different amounts for each

12-month cycle. The ACF for de-trended time series now shows a slight seasonality pattern. The increased observable seasonality is confirmed by observing the changes between the PACF plots: there are significantly increased PACF values as compared to the original time series, for example when lag = 12 and 24.

The ACF and PACF for the hydroelectric plots did not change significantly.

Seasonal Component

Set aside the detrended series and consider the original series again from Q1 to answer Q6 to Q8.

Q6

Just by looking at the time series and the acf plots, do the series seem to have a seasonal trend? No need to run any code to answer your question. Just type in you answer below.

Ans: The time series for renewable generation does not show a significant seasonality, while the time series for hydroelectric consumption shows a significant seasonality due to its sinusoidal pattern.

Q7

Use function *lm()* to fit a seasonal means model (i.e. using the seasonal dummies) the two time series. Ask R to print the summary of the regression. Interpret the regression output. From the results which series have a seasonal trend? Do the results match you answer to Q6?

```r
dummies_renewable <- seasonaldummy(ts_power_data[,1])
dummies_hydro <- seasonaldummy(ts_power_data[,2])

seasonal_means_renewable <- lm(power_data[,1] ~ dummies_renewable)
seasonal_means_hydro <- lm(power_data[,2] ~ dummies_hydro)

summary(seasonal_means_renewable)
```

```
##
## Call:
## lm(formula = power_data[, 1] ~ dummies_renewable)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -205.65  -91.59  -54.59  117.87  356.19
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)           410.598     20.228  20.299   <2e-16 ***
## dummies_renewableJan    1.973     28.468   0.069    0.945
## dummies_renewableFeb  -34.348     28.468  -1.207    0.228
## dummies_renewableMar    4.721     28.468   0.166    0.868
## dummies_renewableApr   -7.896     28.468  -0.277    0.782
## dummies_renewableMay    7.199     28.468   0.253    0.800
## dummies_renewableJun   -3.394     28.468  -0.119    0.905
## dummies_renewableJul    2.850     28.468   0.100    0.920
## dummies_renewableAug   -4.430     28.468  -0.156    0.876
## dummies_renewableSep  -29.037     28.468  -1.020    0.308
```

```
## dummies_renewableOct  -20.205     28.606  -0.706    0.480
## dummies_renewableNov  -20.706     28.606  -0.724    0.469
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 144.5 on 609 degrees of freedom
## Multiple R-squared:  0.008696,   Adjusted R-squared:  -0.009209
## F-statistic: 0.4857 on 11 and 609 DF,  p-value: 0.9126
```

```r
summary(seasonal_means_hydro)
```

```
##
## Call:
## lm(formula = power_data[, 2] ~ dummies_hydro)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -31.101  -6.241  -0.444   6.410  32.363
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)         79.981      1.452  55.094  < 2e-16 ***
## dummies_hydroJan     4.941      2.043   2.418 0.015880 *
## dummies_hydroFeb    -2.513      2.043  -1.230 0.219219
## dummies_hydroMar     7.053      2.043   3.452 0.000595 ***
## dummies_hydroApr     5.399      2.043   2.642 0.008441 **
## dummies_hydroMay    13.907      2.043   6.807 2.40e-11 ***
## dummies_hydroJun    10.729      2.043   5.251 2.09e-07 ***
## dummies_hydroJul     4.033      2.043   1.974 0.048845 *
## dummies_hydroAug    -5.399      2.043  -2.643 0.008436 **
## dummies_hydroSep   -16.596      2.043  -8.123 2.54e-15 ***
## dummies_hydroOct   -16.370      2.053  -7.973 7.66e-15 ***
## dummies_hydroNov   -10.805      2.053  -5.263 1.97e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.37 on 609 degrees of freedom
## Multiple R-squared:  0.4695, Adjusted R-squared:  0.4599
## F-statistic:    49 on 11 and 609 DF,  p-value: < 2.2e-16
```

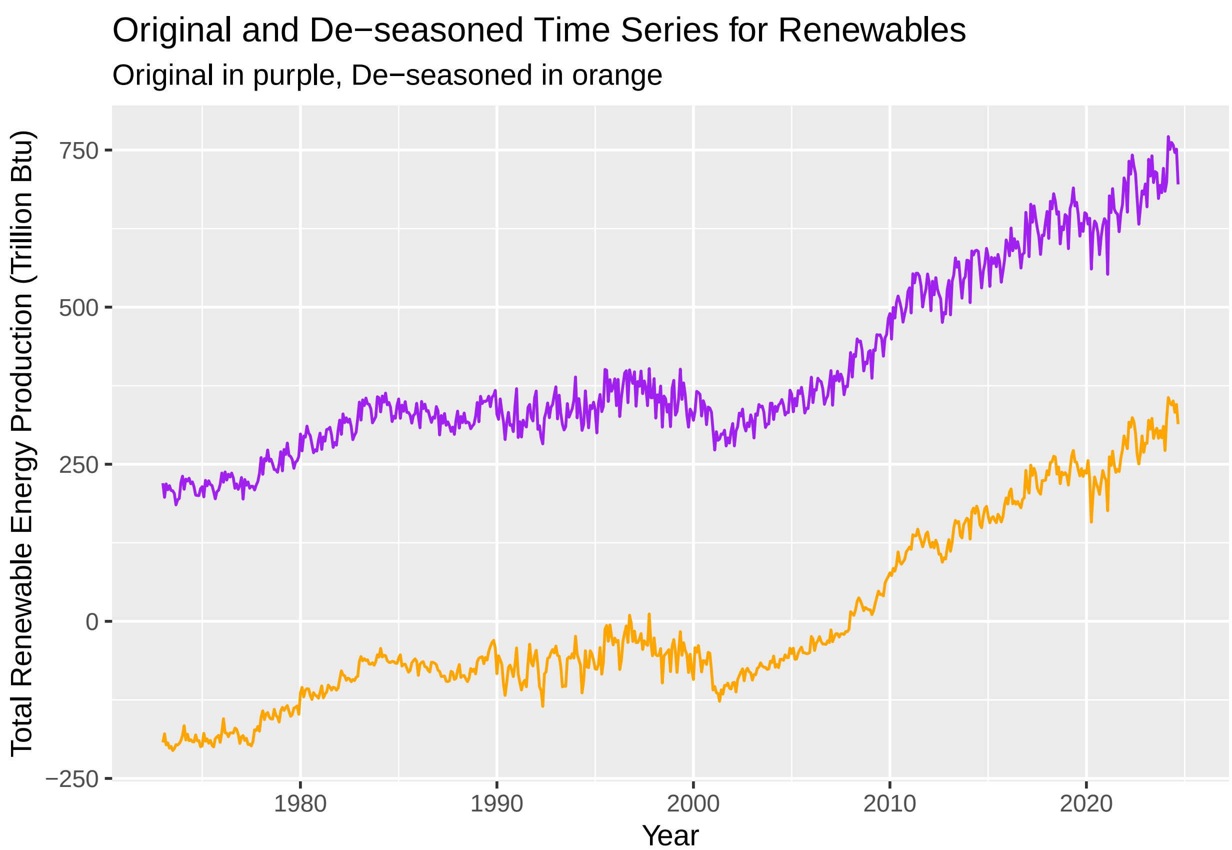
Ans: From the p-value of dummies, the seasonality is only significant for renewable energy production in Decembers, and not significant in other months. The seasonality is significant under 95% confidence for all months except in February for hydroelectric power consumption. These results matches with the observations in Q6.

Q8

Use the regression coefficients from Q7 to deseason the series. Plot the deseason series and compare with the plots from part Q1. Did anything change?

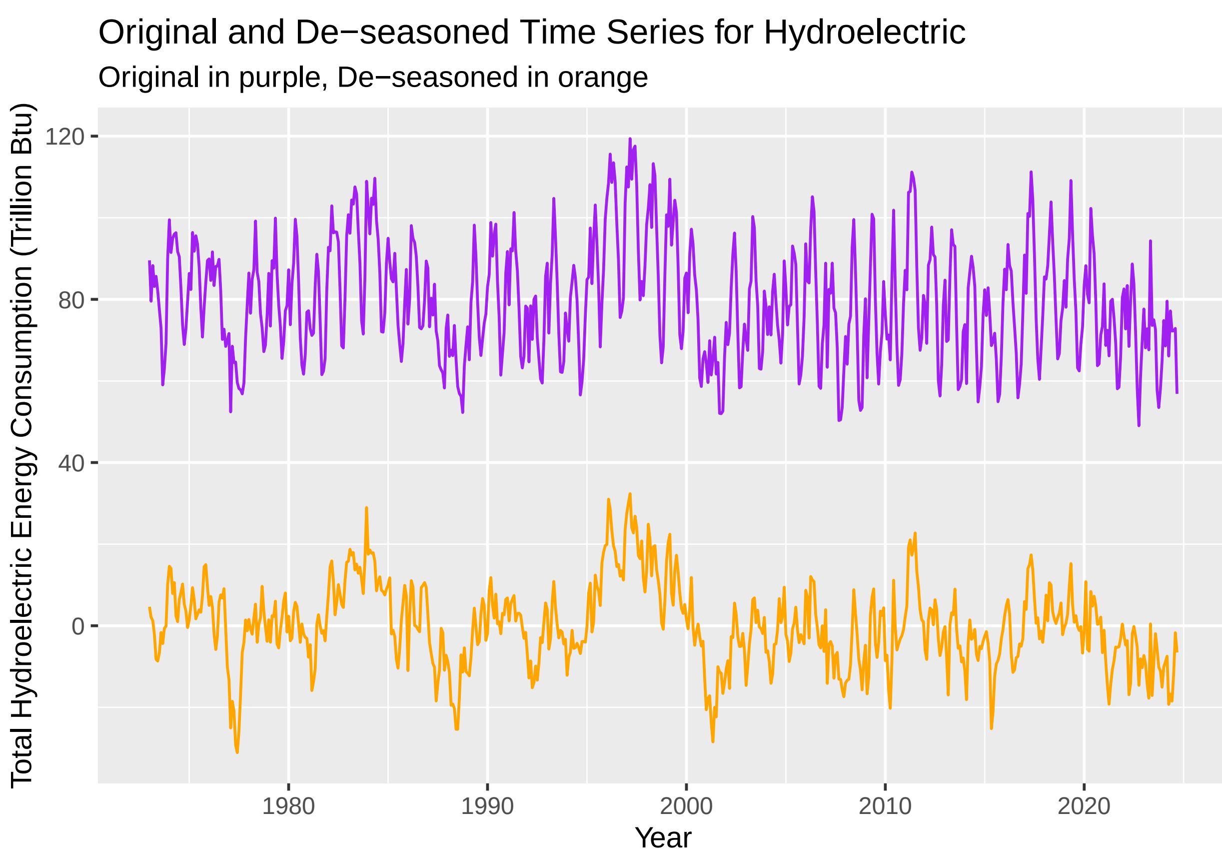
```r
# Renewable
beta_intercept_renewable <-seasonal_means_renewable$coefficients[1]
```

```
beta_coeff_renewable <-seasonal_means_renewable$coefficients[2:12]

renewable_seasonal_component <- array(0,nobs)
for(i in 1:nobs){
  renewable_seasonal_component[i] <-
    beta_intercept_renewable + beta_coeff_renewable %*% dummies_renewable[i,]
}

deseason_renewable <- ts_power_data[,1] - renewable_seasonal_component

ts_deseason_renewable <- ts(deseason_renewable,start=c(1973,1),
                              frequency = 12)

autoplot(ts_power_data[,1], color="purple")+
  autolayer(ts_deseason_renewable,color="orange")+
  ggtitle(
    "Original and De-seasoned Time Series for Renewables",
    subtitle = "Original in purple, De-seasoned in orange") +
  ylab("Total Renewable Energy Production (Trillion Btu)") + xlab("Year")
```




```
# Hydro
beta_intercept_hydro <-seasonal_means_hydro$coefficients[1]
beta_coeff_hydro <-seasonal_means_hydro$coefficients[2:12]

hydro_seasonal_component <- array(0,nobs)
```

```r
for(i in 1:nobs){
  hydro_seasonal_component[i] <-
    beta_intercept_hydro + beta_coeff_hydro %*% dummies_hydro[i,]
}

deseason_hydro <- ts_power_data[,2] - hydro_seasonal_component

ts_deseason_hydro <- ts(deseason_hydro,start=c(1973,1),
                              frequency = 12)

autoplot(ts_power_data[,2], color="purple")+
  autolayer(ts_deseason_hydro,color="orange")+
  ggtitle(
    "Original and De-seasoned Time Series for Hydroelectric",
    subtitle = "Original in purple, De-seasoned in orange") +
  ylab("Total Hydroelectric Energy Consumption (Trillion Btu)") + xlab("Year")
```

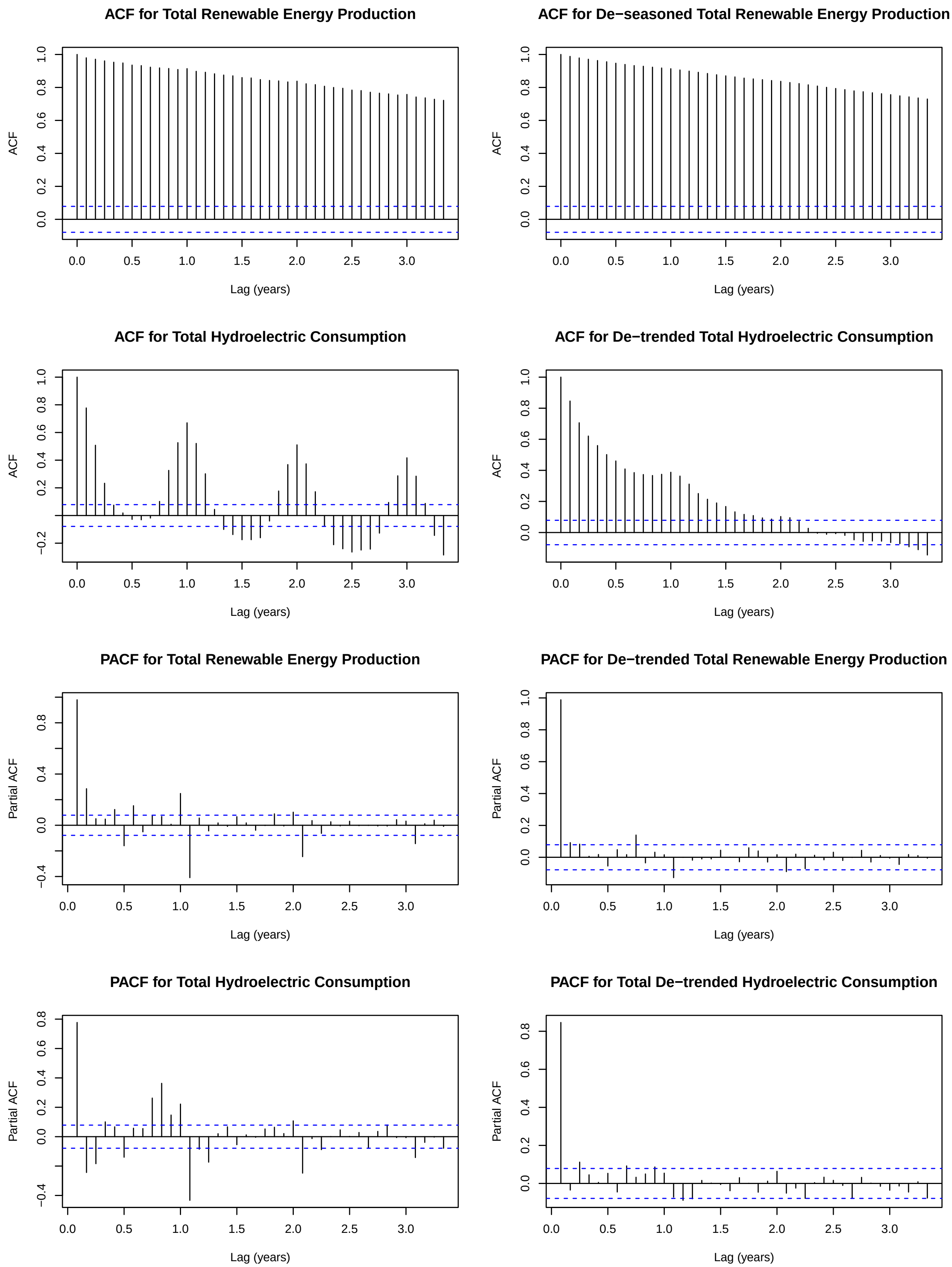


Ans: For renewable energy production, there is a slight change in decrease in the fluctuations, especially between 2000 and 2010, but generally the pattern stays the same. However, for hydroelectric consumption, the seasonal fluctuations decreased significantly, as there are less ups and downs.

Q9

Plot ACF and PACF for the deseason series and compare with the plots from Q1. You may use plot_grid() again to get them side by side, but not mandatory. Did the plots change? How?

```r
par(mfrow=c(4,2))
# ACF plot
renewable_acf<-acf(ts_power_data[,1],lag.max=40,
                  main = "ACF for Total Renewable Energy Production",
                  type="correlation", plot=TRUE,
                  xlab="Lag (years)")

renewable_deseason_acf<-acf(ts_deseason_renewable,lag.max=40,
                  main = "ACF for De-seasoned Total Renewable Energy Production",
                  type="correlation", plot=TRUE,
                  xlab="Lag (years)")

hydro_acf<-acf(ts_power_data[,2],lag.max=40,
              main = "ACF for Total Hydroelectric Consumption",
              type="correlation", plot=TRUE,
              xlab="Lag (years)")

hydro_deseasoned_acf<-acf(ts_deseason_hydro,lag.max=40,
              main = "ACF for De-trended Total Hydroelectric Consumption",
              type="correlation", plot=TRUE,
              xlab="Lag (years)")

# PACF plot
renewable_pacf<-pacf(ts_power_data[,1],lag.max=40,
                  main = "PACF for Total Renewable Energy Production",
                  plot=TRUE,xlab="Lag (years)")

renewable_deseasoned_pacf<-pacf(ts_deseason_renewable,lag.max=40,
                  main = "PACF for De-trended Total Renewable Energy Production",
                  plot=TRUE,xlab="Lag (years)")

hydro_pacf<-pacf(ts_power_data[,2],lag.max=40,
              main = "PACF for Total Hydroelectric Consumption",
              plot=TRUE,xlab="Lag (years)")

hydro_deseasoned_pacf<-pacf(ts_deseason_hydro,lag.max=40,
              main = "PACF for Total De-trended Hydroelectric Consumption",
              plot=TRUE,xlab="Lag (years)")
```



Ans: The ACF plot for renewable energy production does not change significantly, while the ACF plot for hydroelectric consumption changed drastically, since the season fluctuations are being replaced by a decreasing exponential pattern. This suggests that the time series for renewables does not have a strong seasonal component, while that for hydroelectric does.

The PACF plots for both time series changed, in terms of the decrease in significant PACF values. In the deseasoned PACF plots for both time series, there are only two significant PACF values, while in the original, there are always multiple ones. This suggests that the seasonal components for both have successfully been eliminated.