# ENV 790.30 - Time Series Analysis for Energy Data | Spring 2023

Assignment 3 - Due date 02/10/23

Kelsey Husted

## 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_A02_Sp23.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_Consumption
The data comes from the US Energy Information and Administration and corresponds to the December 2022 **Monthly** Energy Review. Once again you will work only with the following columns: Total Biomass Energy Production, Total Renewable Energy Production, Hydroelectric Power Consumption. Create a data frame structure with these three time series only.

```
#Importing data set
energy_data <- read.csv(file = "./Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source


#select specified columns; will include the time column to practice reformatting date
energy_data <-energy_data %>%
  select(Month, Total.Biomass.Energy.Consumption, Total.Renewable.Energy.Production, Hydroelectric.Powe

#Transform date column
Date <- ym(energy_data[,1])
#reintroduce transformed date column to data set
energy_data <- cbind(Date, energy_data[,2:4])
head(energy_data)
```

```
##         Date Total.Biomass.Energy.Consumption Total.Renewable.Energy.Production
## 1 1973-01-01                          129.787                           403.981
## 2 1973-02-01                          117.338                           360.900
```

```
## 3 1973-03-01                        129.938                            400.161
## 4 1973-04-01                        125.636                            380.470
## 5 1973-05-01                        129.834                            392.141
## 6 1973-06-01                        125.611                            377.232
##   Hydroelectric.Power.Consumption
## 1                         272.703
## 2                         242.199
## 3                         268.810
## 4                         253.185
## 5                         260.770
## 6                         249.859
```

```
#ts transformation
ts_energy_data <- ts(energy_data[,2:4], frequency = 12, start = c(1973,1))
```

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
library(tseries)
require(lubridate)
library(forecast)
library(dplyr)
library(Kendall)
library(ggplot2)
library(tidyverse)
```

##Trend Component

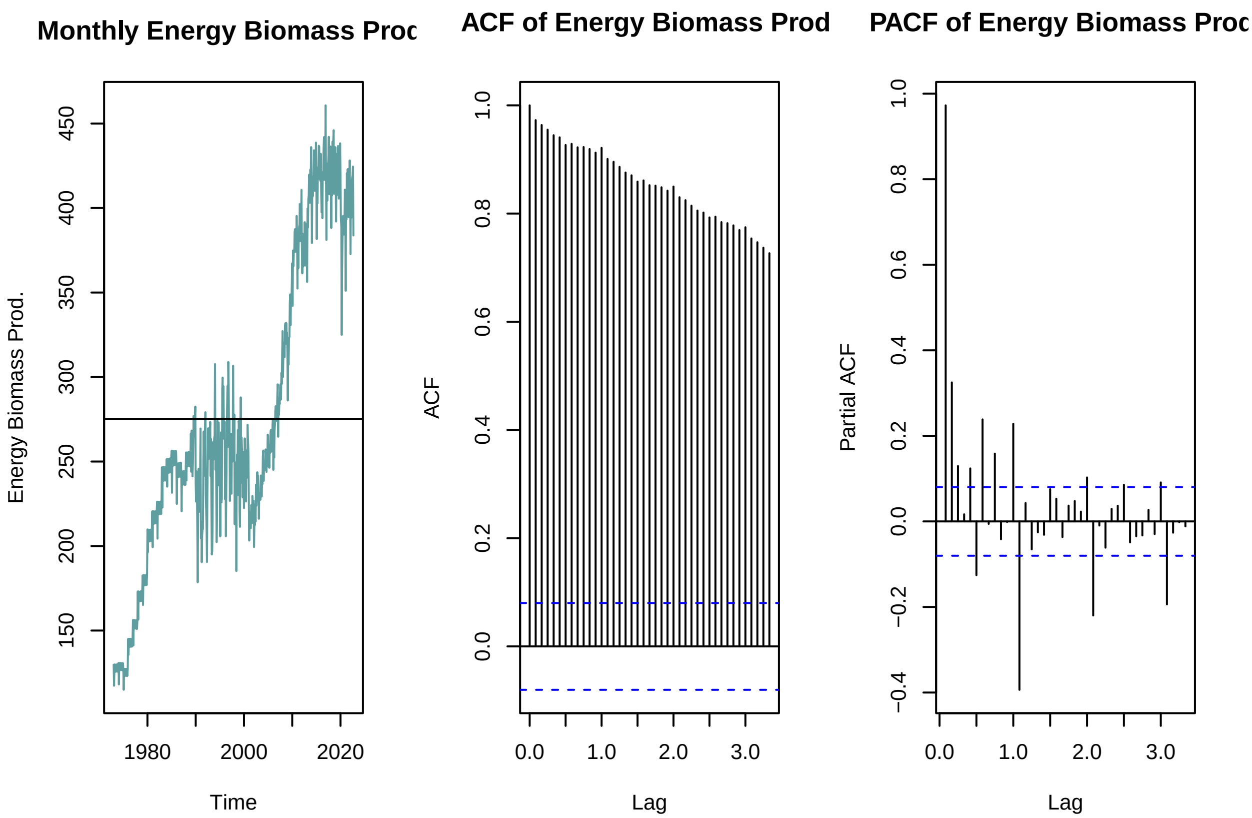
### Q1

Create a plot window that has one row and three columns. And then for each object on your data frame, fill the plot window with time series plot, ACF and PACF. You may use the some code form A2, but I want all three plots on the same window this time. (Hint: use par() function)

```
#Total Biomass Energy Production
#meant for plot#1 scripted below
Biomass_mean <- mean(ts_energy_data[,1])

par(mfrow = c(1,3))
#Plot 1
plot(ts_energy_data[,1],main = "Monthly Energy Biomass Prod.",
         col = "cadetblue",
         ylab = "Energy Biomass Prod.") +
         abline(h = Biomass_mean)
```

```
## integer(0)
```

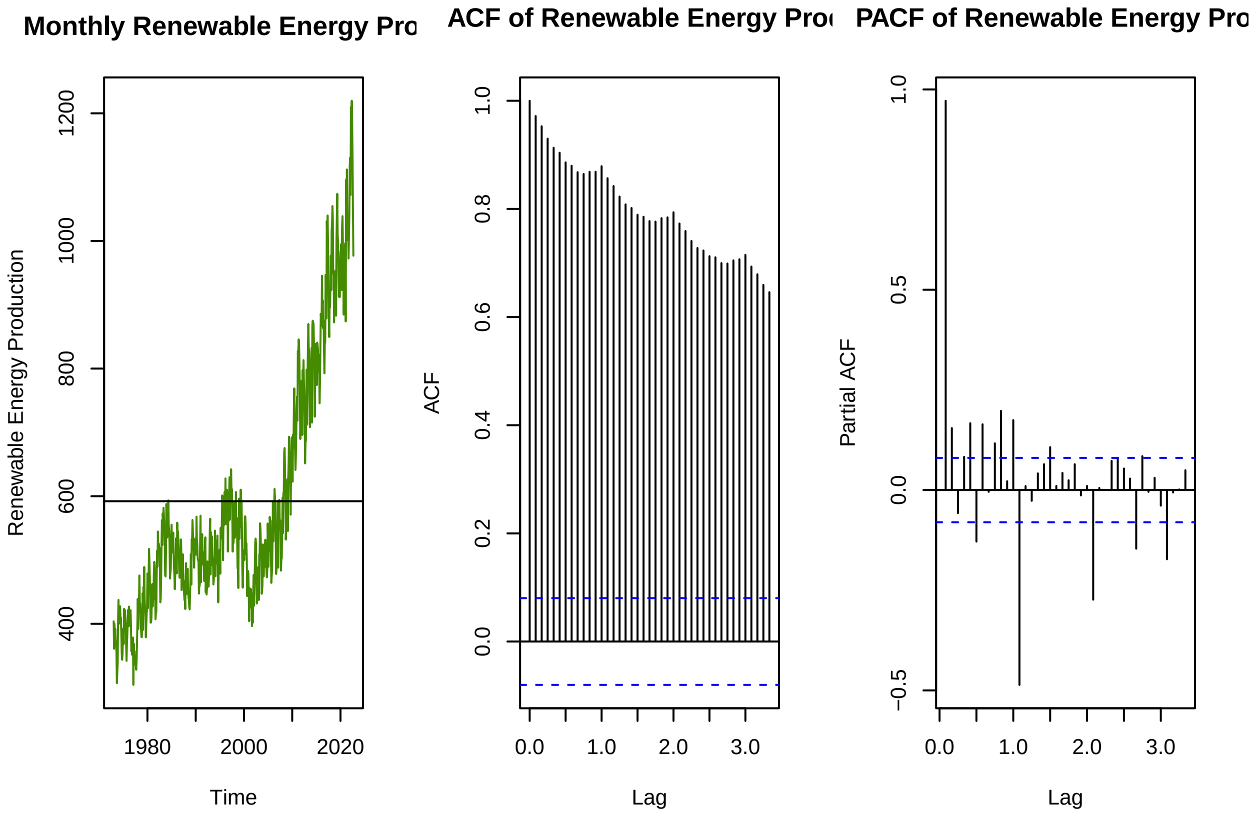
```
#Plot 2
acf(ts_energy_data[,1],lag.max=40, type="correlation",plot=TRUE, main = "ACF of Energy Biomass Prod.")
#Plot 3
pacf(ts_energy_data[,1],lag.max=40, plot=TRUE,  main = "PACF of Energy Biomass Prod.")
```




```
#Total Renewable Energy Production
#meant for plot#1 scripted below
Renewable_mean <- mean(ts_energy_data[,2])

par(mfrow = c(1,3))
#Plot 1
plot(ts_energy_data[,2],main = "Monthly Renewable Energy Prod.",
         col = "chartreuse4",
         ylab = "Renewable Energy Production") +
         abline(h = Renewable_mean)
```

```
## integer(0)
```

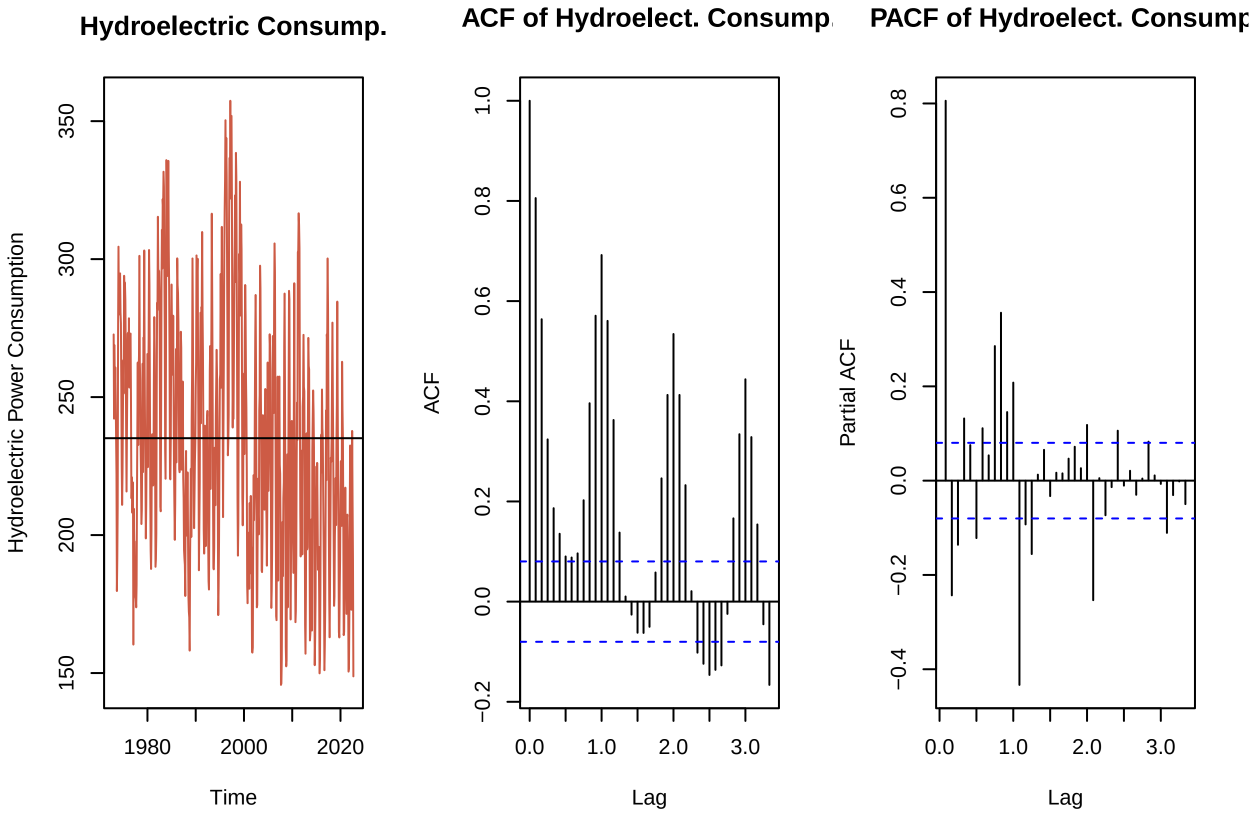
```
#Plot 2
acf(ts_energy_data[,2],lag.max=40, type="correlation",plot=TRUE, main = "ACF of Renewable Energy Prod.")
#Plot 3
pacf(ts_energy_data[,2],lag.max=40, plot=TRUE,  main = "PACF of Renewable Energy Prod.")
```




```r
#Hydroelectric Power Consumption
#meant for plot#1 scripted below
Hydroelect_mean <- mean(ts_energy_data[,3])

par(mfrow = c(1,3))
#Plot 1
plot(ts_energy_data[,3],main = "Hydroelectric Consump.",
         col = "coral3",
         ylab = "Hydroelectric Power Consumption") +
         abline(h = Hydroelect_mean)
```

```
## integer(0)
```

```r
#Plot 2
acf(ts_energy_data[,3],lag.max=40, type="correlation",plot=TRUE, main = "ACF of Hydroelect. Consump.")
#Plot 3
pacf(ts_energy_data[,3],lag.max=40, plot=TRUE,  main = "PACF of Hydroelect. Consump.")
```




**Q2**

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

> Answer: According to the Q1 plots of the Total Biomass Energy Production and Total Renewable Energy Production, there appears to be a trend since autocorrelations for small lags tend to be large and positive which is generally associated with a trend when analyzing ACF plots. The slow decrease in the ACF as the lags increase is commonly associated with trends.The trend for both Biomass and Renewable Energy appears to be positive with the uphill trend displayed in the general data plots. Alternatively, the Hydroelectric Consumption plots from Q1 looks like there may be a slight downward trend present with seasonality. More analysis will definitely be needed to assess the potential significance of this trend for Hydroelectric Consumption.

**Q3**

Use the *lm()* function to fit a linear trend to the three 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.

```
#Total Biomass Energy Production

num.obs <- nrow(energy_data)
observ <- 1:num.obs

#Fit a linear trend to TS
linear_biomass <- lm(energy_data[,2] ~ observ)
summary(linear_biomass)
```

```
##
## Call:
## lm(formula = energy_data[, 2] ~ observ)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -99.441 -24.384   4.598  31.945  79.606
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.369e+02  3.132e+00   43.71   <2e-16 ***
## observ      4.626e-01  9.075e-03   50.98   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 38.21 on 595 degrees of freedom
## Multiple R-squared:  0.8137, Adjusted R-squared:  0.8134
## F-statistic:  2599 on 1 and 595 DF,  p-value: < 2.2e-16
```

```
#The Y-intercept = 133.74 while the slope is 0.480 which indicates a positive relationship between the
#save regression coefficients
beta0_biomass=as.numeric(linear_biomass$coefficients[1])  #first coefficient is the intercept term or b
beta1_biomass=as.numeric(linear_biomass$coefficients[2]) #slope
```

```
#Total Renewable Energy Production
#Fit a linear trend to TS
linear_renewable<- lm(energy_data[,3] ~ observ)
summary(linear_renewable)
```

```
##
## Call:
## lm(formula = energy_data[, 3] ~ observ)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -238.75  -61.85    8.59   64.48  352.27
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 312.2475     8.4902   36.78   <2e-16 ***
## observ        0.9362     0.0246   38.05   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 103.6 on 595 degrees of freedom
## Multiple R-squared:  0.7088, Adjusted R-squared:  0.7083
## F-statistic:  1448 on 1 and 595 DF,  p-value: < 2.2e-16
```

```
#The Y-intercept = 312.24 while the slope is 0.9362 which indicates a positive relationship between the
#save regression coefficients
beta0_renewable=as.numeric(linear_renewable$coefficients[1])  #first coefficient is the intercept term
beta1_renewable=as.numeric(linear_renewable$coefficients[2]) #slope
```

```
#Hydroelectric Power Consumption
#Fit a linear trend to TS
linear_hydro <- lm(energy_data[,4] ~ observ)
```

```
summary(linear_hydro)
```

```
##
## Call:
## lm(formula = energy_data[, 4] ~ observ)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -95.42 -31.20  -2.56  27.32 121.61
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 259.898013   3.427300  75.832  < 2e-16 ***
## observ       -0.082888   0.009931  -8.346 4.94e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 41.82 on 595 degrees of freedom
## Multiple R-squared:  0.1048, Adjusted R-squared:  0.1033
## F-statistic: 69.66 on 1 and 595 DF,  p-value: 4.937e-16
```

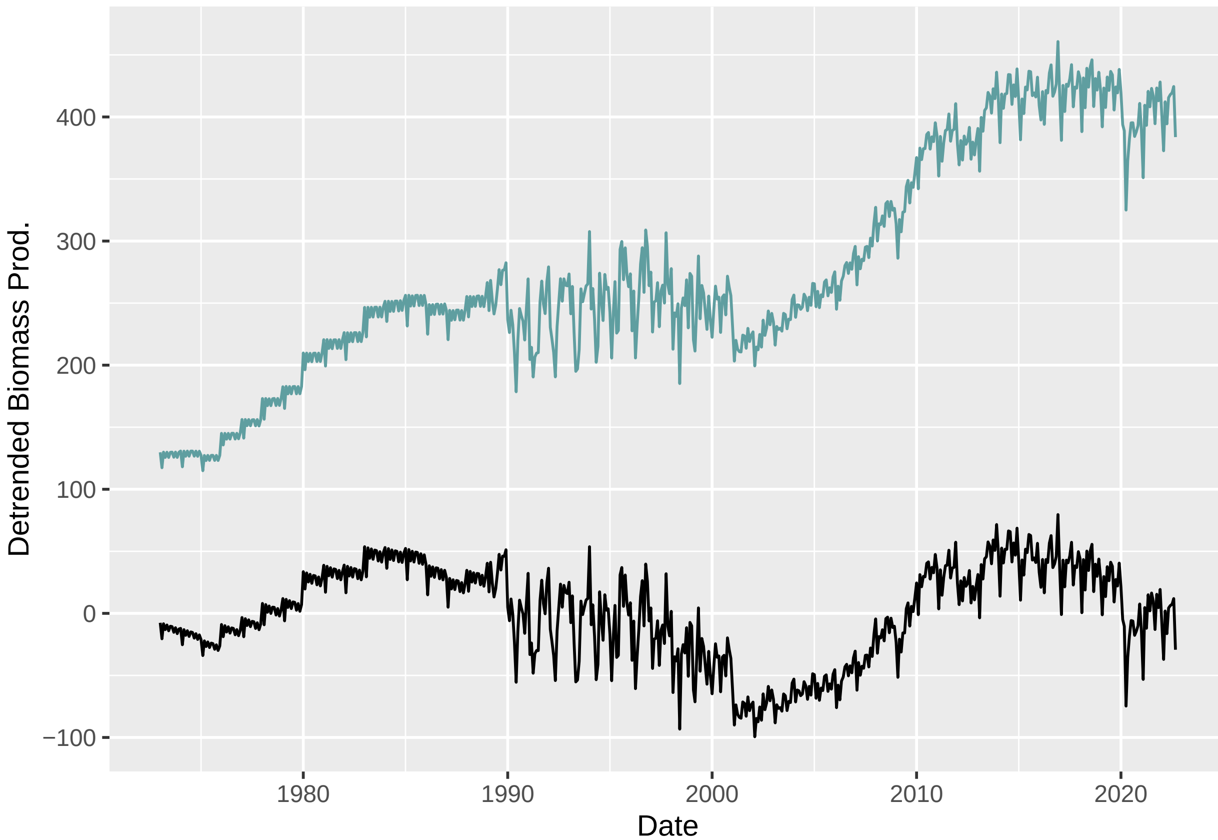
```
#The Y-intercept = 259.89 while the slope is -0.082888 which indicates a slight negative relationship. 
beta0_hydro=as.numeric(linear_hydro$coefficients[1])  #first coefficient is the intercept term or beta0
beta1_hydro=as.numeric(linear_hydro$coefficients[2]) #slope
```

### Q4

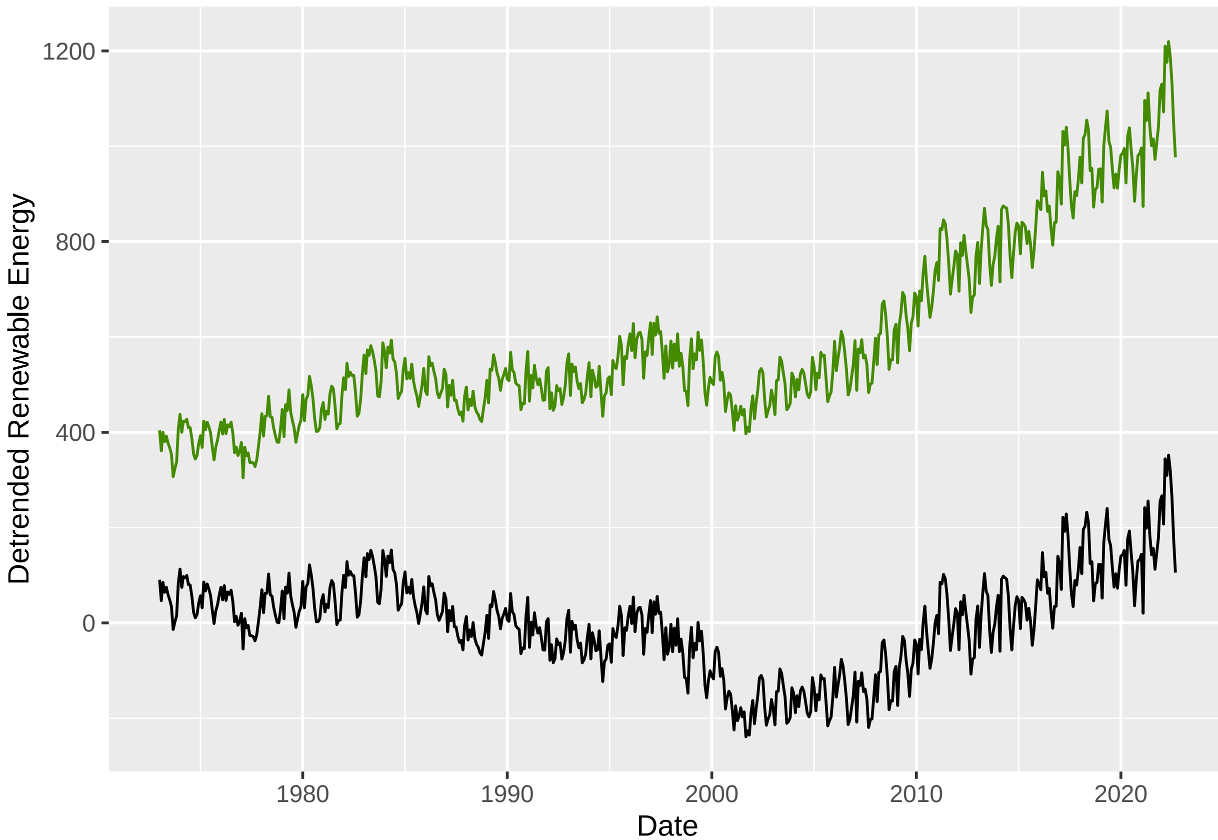
Use the regression coefficients from Q3 to detrend the series. Plot the detrended series and compare with the plots from Q1. What happened? Did anything change?

> Answer: For all graphs the trends were removed that were once noticeable in the previous graphs created in Q1. For instance, the Biomass and Renewable Energy plots don't show the consistent upward trend after being detrended and the Hydroelectric plot doesn't display the slight downward trend after being detrended.

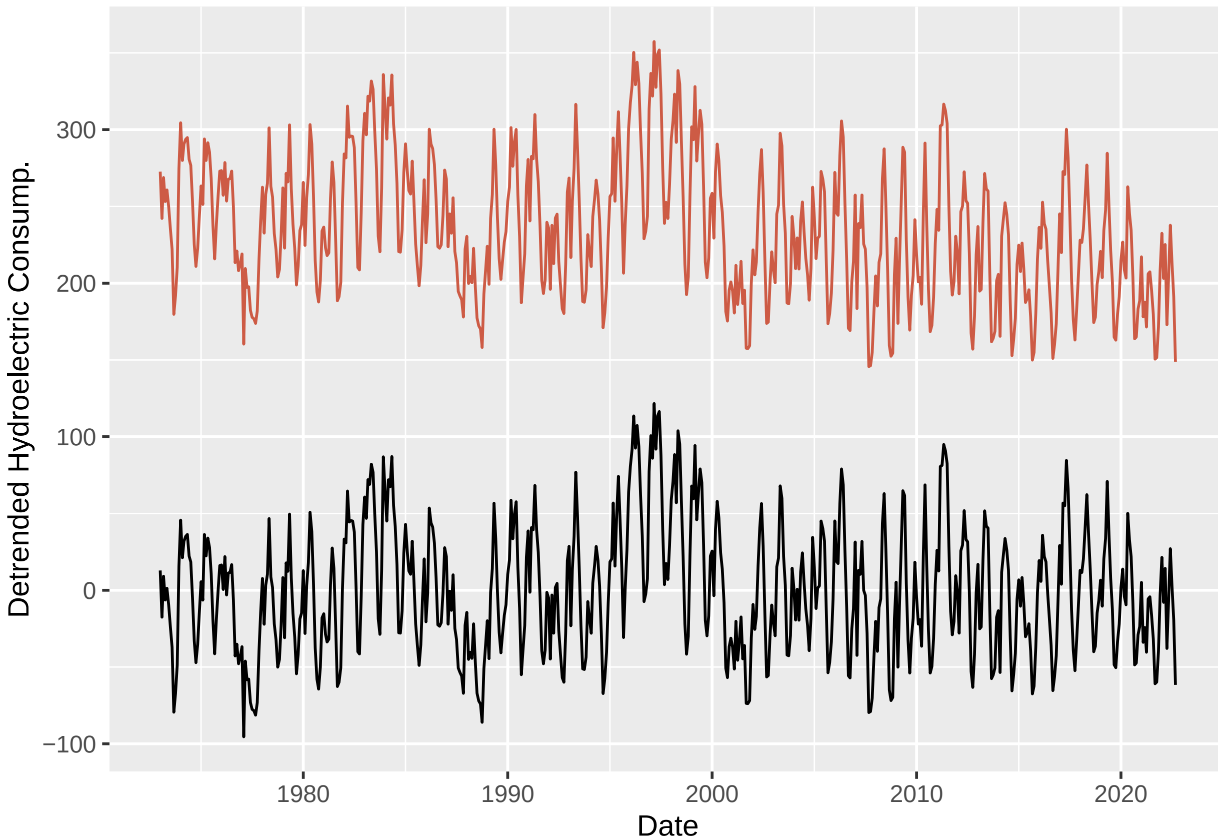
```
par(mfrow = c(1,3))
#Total Biomass Energy Production
Biomass_detrend <- energy_data[,2]-(beta0_biomass+beta1_biomass*observ)

ggplot(energy_data, aes(x = Date, y = energy_data[,2])) +
  geom_line(color="cadetblue") +
  ylab(paste0("Detrended Biomass Prod.")) +
  geom_line(aes(y=Biomass_detrend))
```




```
#Total Renewable Energy Production
RenewableEnergy_detrend <- energy_data[,3]-(beta0_renewable+beta1_renewable*observ)

ggplot(energy_data, aes(x = Date, y = energy_data[,3])) +
  geom_line(color="chartreuse4") +
  ylab(paste0("Detrended Renewable Energy")) +
  geom_line(aes(y=RenewableEnergy_detrend))
```




```r
#Hydroelectric Power Consumption
hydroelectric_detrend <- energy_data[,4]-(beta0_hydro+beta1_hydro*observ)

ggplot(energy_data, aes(x = Date, y = energy_data[,4])) +
  geom_line(color="coral3") +
  ylab(paste0("Detrended Hydroelectric Consump.")) +
  geom_line(aes(y=hydroelectric_detrend))
```

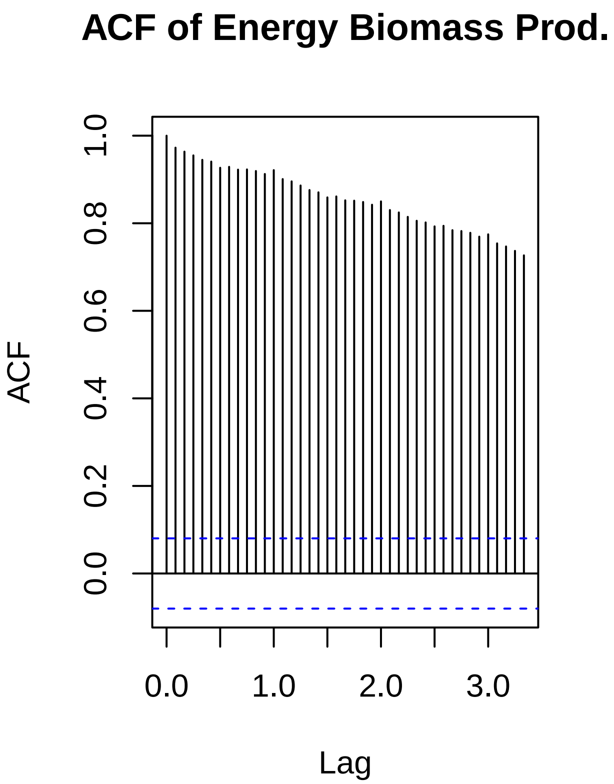


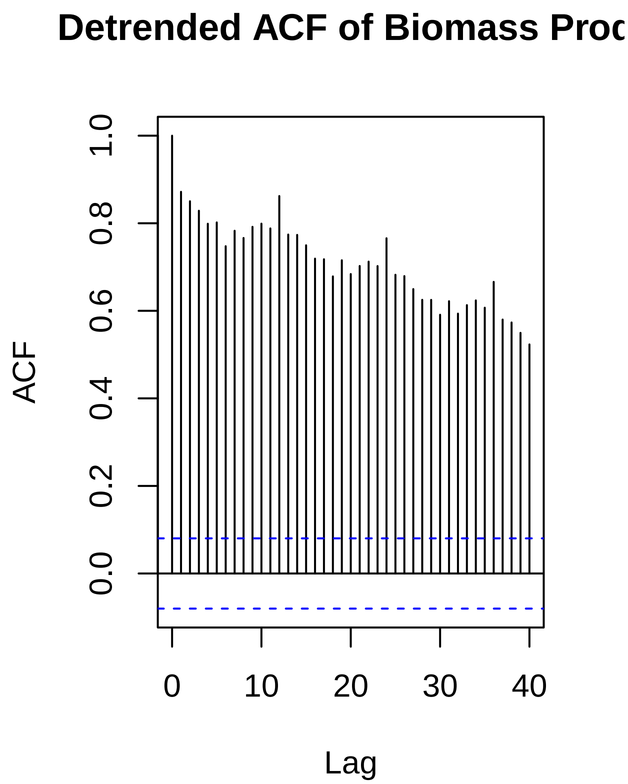

**Q5**

Plot ACF and PACF for the detrended series and compare with the plots from Q1. Did the plots change? How?

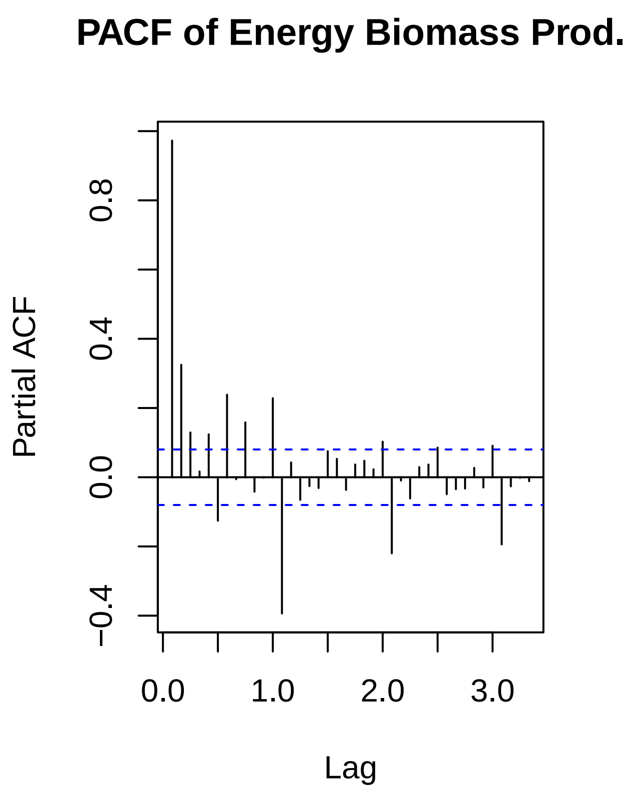
> Answer: I didn't notice much change between the previous acf and pacf plots created in Q1 compared to the detrended versions displayed below.

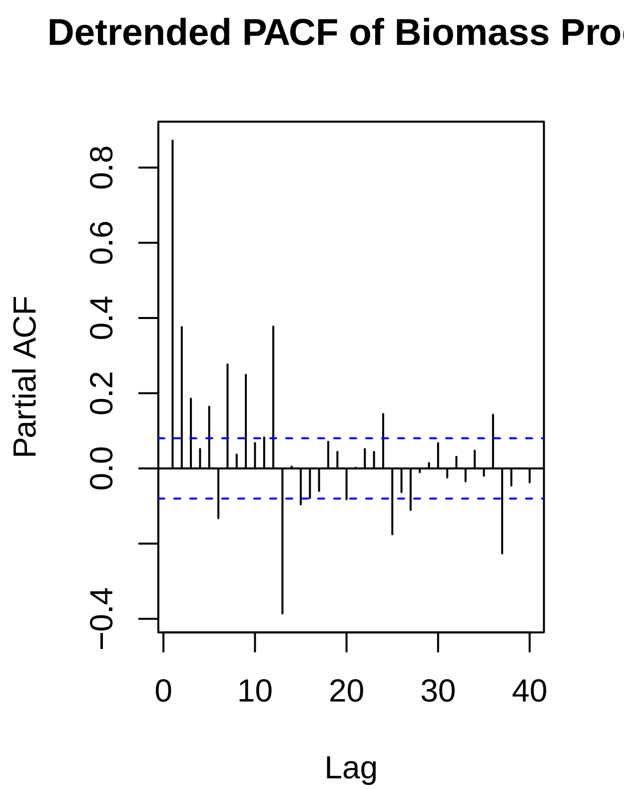
```
par(mfrow = c(1,2))


#Total Biomass Energy Production
acf(ts_energy_data[,1],lag.max=40, type="correlation",plot=TRUE, main = "ACF of Energy Biomass Prod.")
acf(Biomass_detrend,lag.max=40, plot=TRUE, main = "Detrended ACF of Biomass Prod.")
```



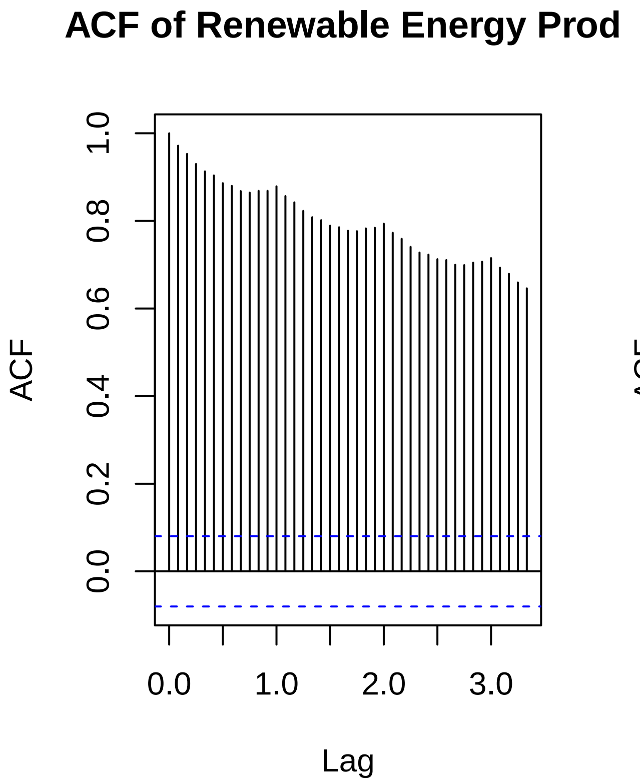

```
pacf(ts_energy_data[,1],lag.max=40, plot=TRUE,  main = "PACF of Energy Biomass Prod.")
pacf(Biomass_detrend,lag.max=40, plot=TRUE, main = "Detrended PACF of Biomass Prod.")
```

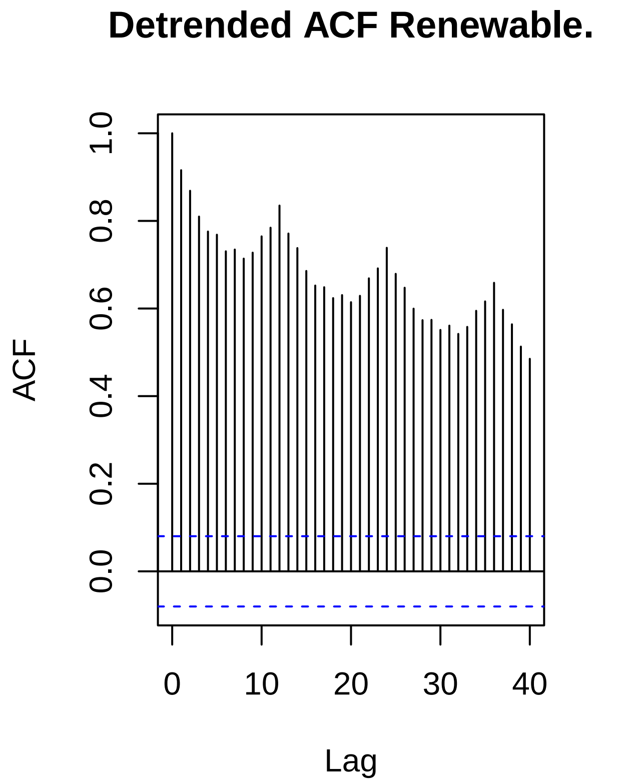


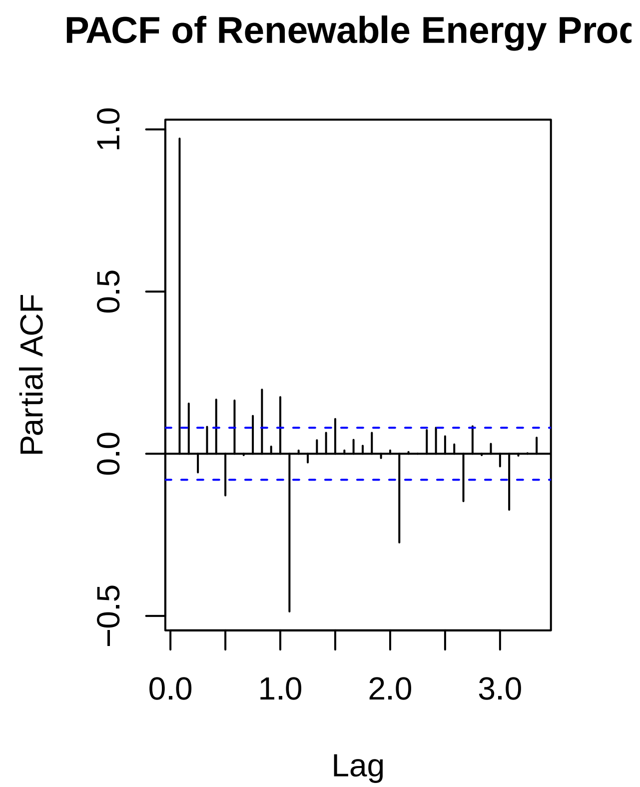

```
#Total Renewable Energy Production
acf(ts_energy_data[,2],lag.max=40, type="correlation",plot=TRUE, main = "ACF of Renewable Energy Prod."
```

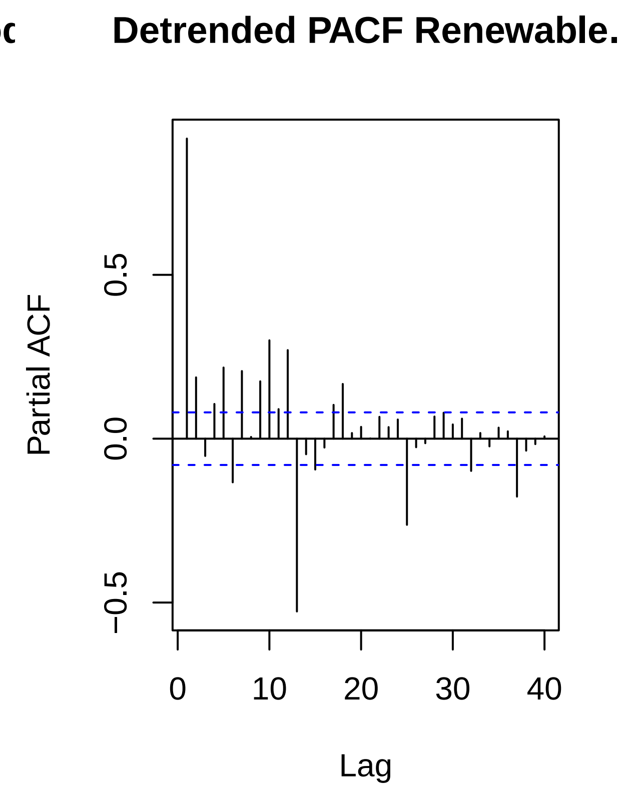
```r
acf(RenewableEnergy_detrend,lag.max=40,plot=TRUE, main = "Detrended ACF Renewable.")
```



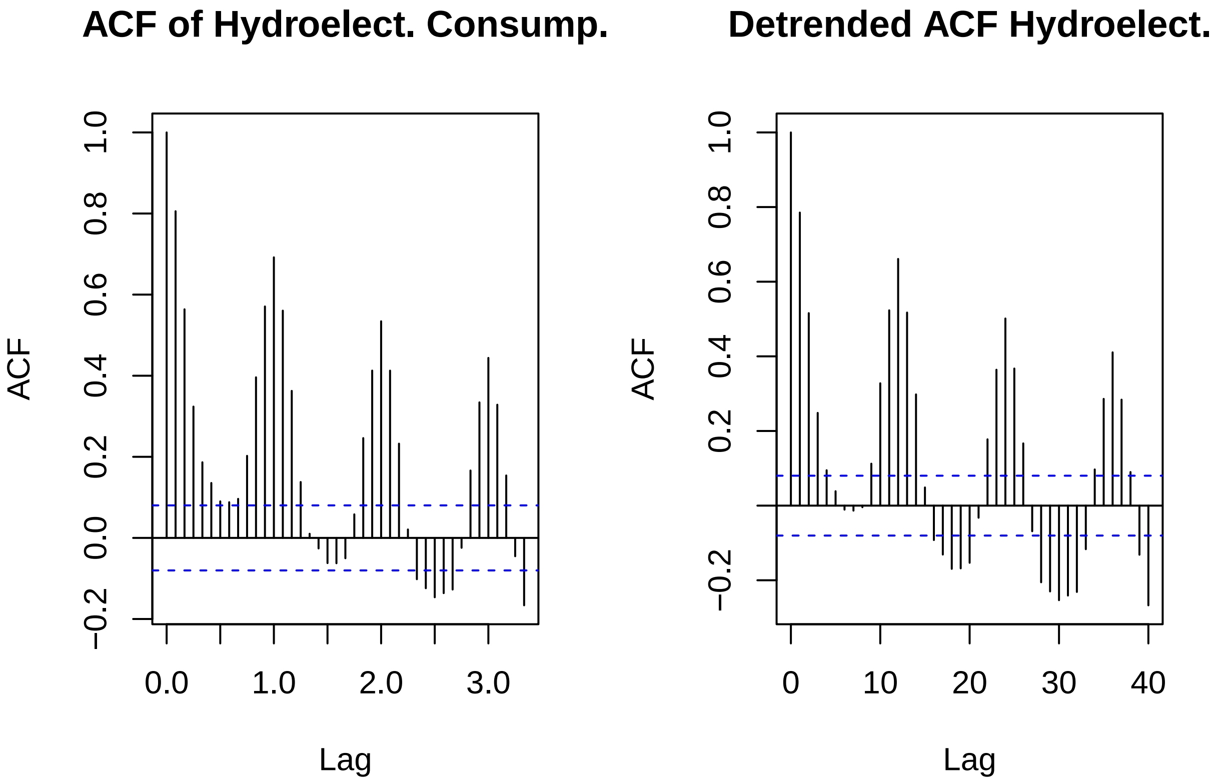




```r
pacf(ts_energy_data[,2],lag.max=40, plot=TRUE,  main = "PACF of Renewable Energy Prod.")
pacf(RenewableEnergy_detrend,lag.max=40,plot=TRUE, main = "Detrended PACF Renewable.")
```

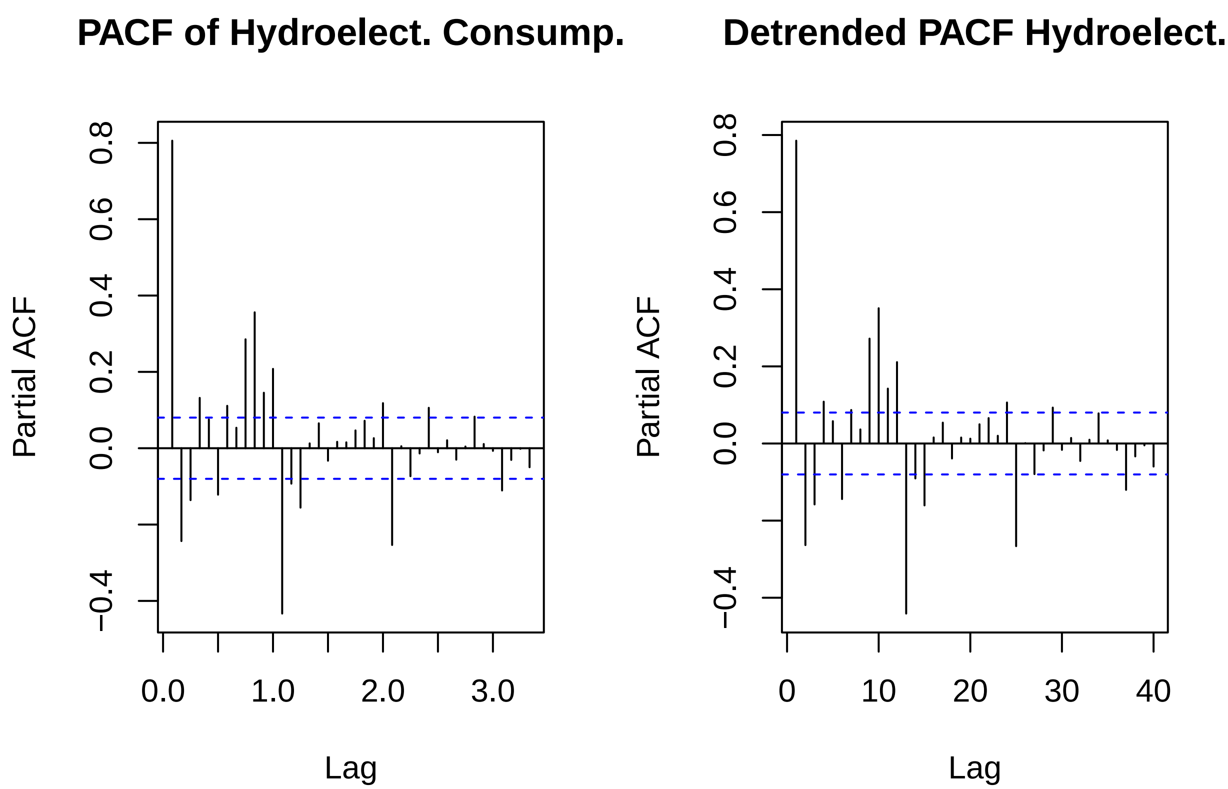


```r
#Hydroelectric Power Consumption
acf(ts_energy_data[,3],lag.max=40, type="correlation",plot=TRUE, main = "ACF of Hydroelect. Consump.")
acf(hydroelectric_detrend,lag.max=40, plot=TRUE, main = "Detrended ACF Hydroelect.")
```




```r
pacf(ts_energy_data[,3],lag.max=40, plot=TRUE,  main = "PACF of Hydroelect. Consump.")
pacf(hydroelectric_detrend,lag.max=40, plot=TRUE, main = "Detrended PACF Hydroelect.")
```




## Seasonal Component

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

### Q6

Do the series seem to have a seasonal trend? Which serie/series? Use function *lm()* to fit a seasonal means model (i.e. using the seasonal dummies) to this/these time series. Ask R to print the summary of the regression. Interpret the regression output. Save the regression coefficients for further analysis.

> Answer: When running the seasonal means model, the Total Biomass Production and Renewable Energy Production both displayed p-values $> 0.05$ so the null hypothesis is accepted and seasonality is not present since the correlation is not significant. Alternatively, the seasonal means model for Hydroelectic Consumption had a p-value $< 0.05$ so the null hypothesis is rejected and seasonality is thus present.

```
#Create dummies for three TS
dummies <- seasonaldummy(ts_energy_data[,1])

#Fit linear model to seasonal dummies
#Total Biomass Production
seas_biomass_model <- lm(energy_data[,2]~dummies)
summary(seas_biomass_model) #p-value > 0.05 so we accept the null hypothesis (i.e., no seasonality)
```

```
##
## Call:
## lm(formula = energy_data[, 2] ~ dummies)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -159.29  -52.27  -22.79   91.95  175.26
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  285.495     12.685  22.507   <2e-16 ***
## dummiesJan    -4.242     17.849  -0.238   0.8122
## dummiesFeb   -31.454     17.849  -1.762   0.0786 .
## dummiesMar    -9.781     17.849  -0.548   0.5839
## dummiesApr   -20.871     17.849  -1.169   0.2428
## dummiesMay   -12.792     17.849  -0.717   0.4739
## dummiesJun   -17.978     17.849  -1.007   0.3143
## dummiesJul    -2.789     17.849  -0.156   0.8759
## dummiesAug     1.113     17.849   0.062   0.9503
## dummiesSep   -12.343     17.849  -0.692   0.4895
## dummiesOct    -2.455     17.939  -0.137   0.8912
## dummiesNov    -9.347     17.939  -0.521   0.6025
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 88.79 on 585 degrees of freedom
## Multiple R-squared:  0.01091,    Adjusted R-squared:  -0.00769
## F-statistic: 0.5865 on 11 and 585 DF,  p-value: 0.8406
```

```r
beta_int_biomass <- seas_biomass_model$coefficients[1]
beta_coeff_biomass <- seas_biomass_model$coefficients[2:12]
biomass_seas_comp=array(0,num.obs)
for(i in 1:num.obs){
  biomass_seas_comp[i]=(beta_int_biomass+beta_coeff_biomass%*%dummies[i,])
}

#Total Renewable Energy Production
seas_renew_model <- lm(energy_data[,3]~dummies)
summary(seas_renew_model) # p-value > 0.05 so we accept the null hypothesis (i.e., no seasonality)
```

```
##
## Call:
## lm(formula = energy_data[, 3] ~ dummies)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -284.92 -122.23  -68.42   91.22  585.68
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  601.022     27.260  22.048   <2e-16 ***
## dummiesJan    11.468     38.358   0.299    0.765
## dummiesFeb   -41.456     38.358  -1.081    0.280
## dummiesMar    23.130     38.358   0.603    0.547
## dummiesApr     9.959     38.358   0.260    0.795
## dummiesMay    38.853     38.358   1.013    0.312
## dummiesJun    20.378     38.358   0.531    0.595
## dummiesJul     8.298     38.358   0.216    0.829
## dummiesAug   -19.450     38.358  -0.507    0.612
## dummiesSep   -63.770     38.358  -1.662    0.097 .
## dummiesOct   -52.612     38.551  -1.365    0.173
```

```
## dummiesNov   -42.537     38.551  -1.103    0.270
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 190.8 on 585 degrees of freedom
## Multiple R-squared:  0.02844,    Adjusted R-squared:  0.01017
## F-statistic: 1.557 on 11 and 585 DF,  p-value: 0.1076
```

```
beta_int_renew <- seas_renew_model$coefficients[1]
beta_coeff_renew <- seas_renew_model$coefficients[2:12]
renew_seas_comp=array(0,num.obs)
for(i in 1:num.obs){
  renew_seas_comp[i]=(beta_int_renew+beta_coeff_renew%*%dummies[i,])
}

#Hydroelectric Consumption
seas_hydro_model <- lm(energy_data[,4]~dummies)
summary(seas_hydro_model) # p-value < 0.05 so we REJECT the null hypothese (i.e., seasonality present)
```

```
##
## Call:
## lm(formula = energy_data[, 4] ~ dummies)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -88.99 -23.47  -2.81  21.99 100.18
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  237.225      4.878  48.634  < 2e-16 ***
## dummiesJan    13.594      6.864   1.981  0.04811 *
## dummiesFeb    -8.254      6.864  -1.203  0.22964
## dummiesMar    19.980      6.864   2.911  0.00374 **
## dummiesApr    15.649      6.864   2.280  0.02297 *
## dummiesMay    39.210      6.864   5.713 1.77e-08 ***
## dummiesJun    31.209      6.864   4.547 6.61e-06 ***
## dummiesJul    10.436      6.864   1.520  0.12895
## dummiesAug   -17.909      6.864  -2.609  0.00931 **
## dummiesSep   -50.173      6.864  -7.310 8.82e-13 ***
## dummiesOct   -48.262      6.898  -6.996 7.22e-12 ***
## dummiesNov   -32.285      6.898  -4.680 3.56e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 34.14 on 585 degrees of freedom
## Multiple R-squared:  0.4132, Adjusted R-squared:  0.4022
## F-statistic: 37.45 on 11 and 585 DF,  p-value: < 2.2e-16
```

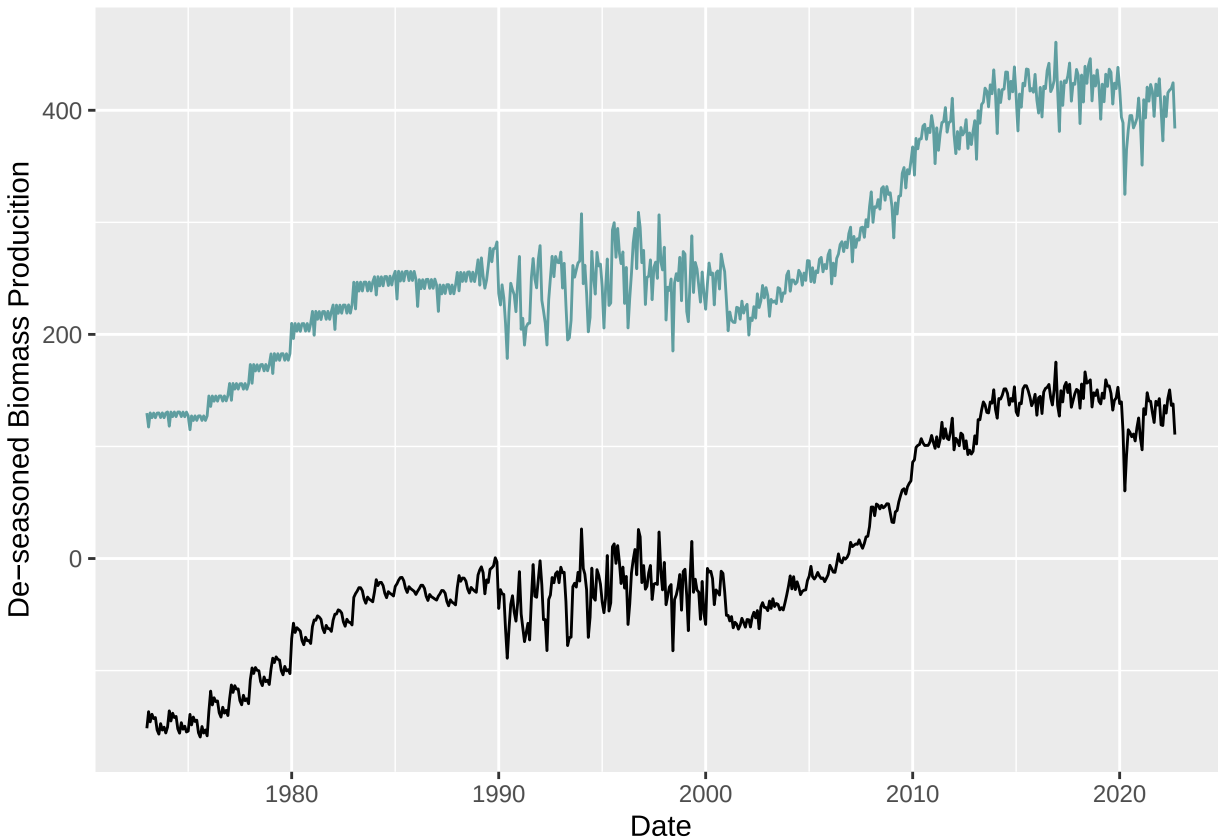
```
beta_int_hydro <- seas_hydro_model$coefficients[1]
beta_coeff_hydro <- seas_hydro_model$coefficients[2:12]
hydro_seas_comp=array(0,num.obs)
for(i in 1:num.obs){
  hydro_seas_comp[i]=(beta_int_hydro+beta_coeff_hydro%*%dummies[i,])
}
```

**Q7**

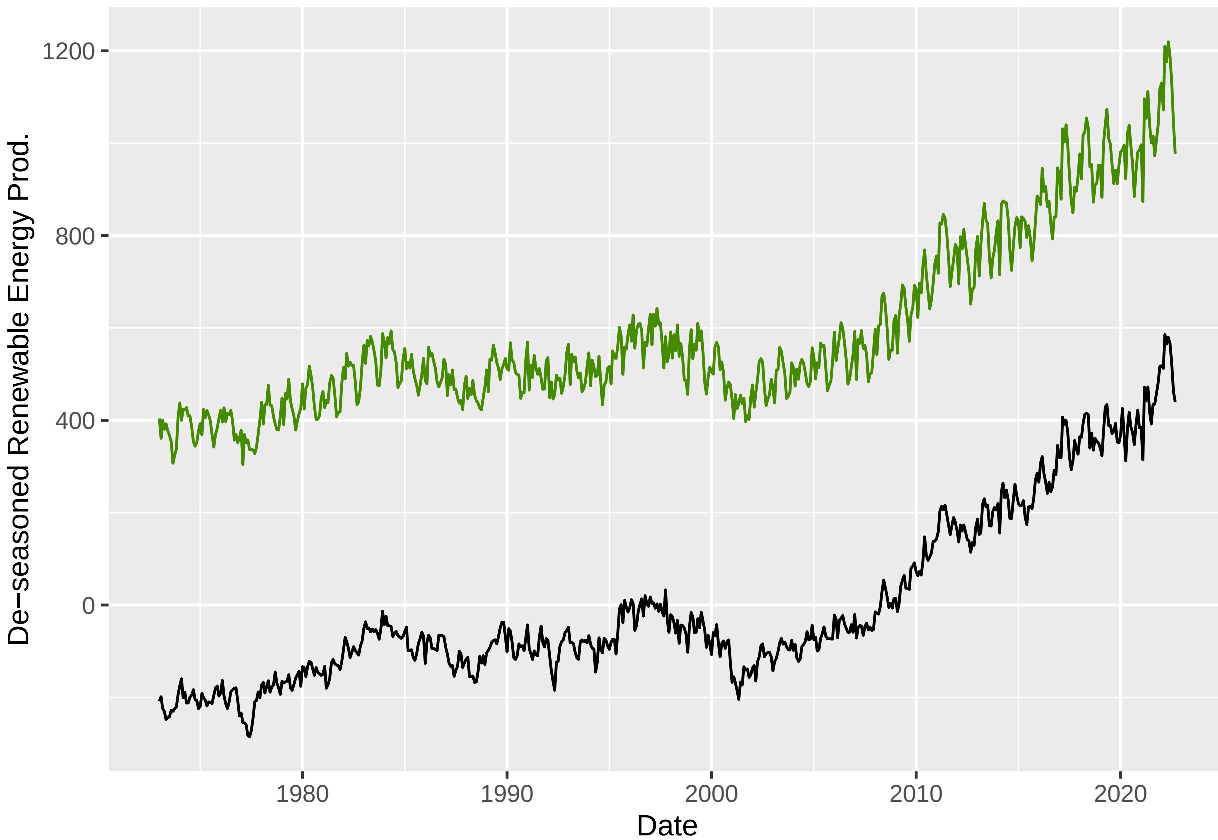
Use the regression coefficients from Q6 to deseason the series. Plot the deseason series and compare with the plots from part Q1. Did anything change?

> Answer: Only the Hydroelectric Consumption data displayed significant seasonality. The de-seasoned graphs displays a smoother version of the Q1 plots will less intense peaks since that seasonal roughness component was removed.

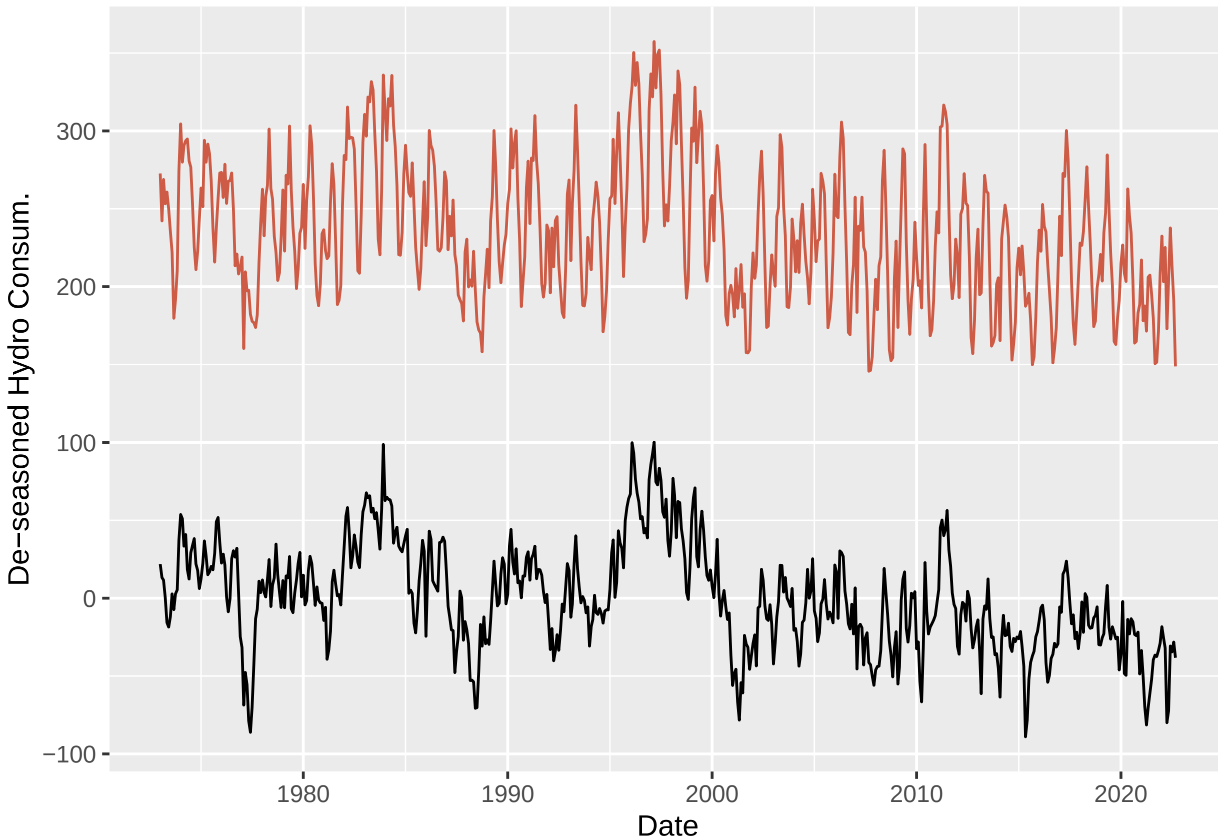
```
par(mfrow=c(1,3))
#Total Biomass Production
deseason_biomass <- energy_data[,2]-biomass_seas_comp
ggplot(energy_data, aes(x = Date, y = energy_data[,2])) +
  geom_line(color = "cadetblue") +
  ylab(paste0("De-seasoned Biomass Producition")) +
  geom_line(aes(y=deseason_biomass))
```




```
#Total Renewable Energy Production
deseason_renew <- energy_data[,3]-renew_seas_comp
ggplot(energy_data, aes(x = Date, y = energy_data[,3])) +
  geom_line(color = "chartreuse4") +
  ylab(paste0("De-seasoned Renewable Energy Prod.")) +
  geom_line(aes(y=deseason_renew))
```




```
#Hydroelectric Consumption
deseason_hydro <- energy_data[,4]-hydro_seas_comp
ggplot(energy_data, aes(x = Date, y = energy_data[,4])) +
  geom_line(color = "coral3") +
  ylab(paste0("De-seasoned Hydro Consum.")) +
  geom_line(aes(y=deseason_hydro))
```

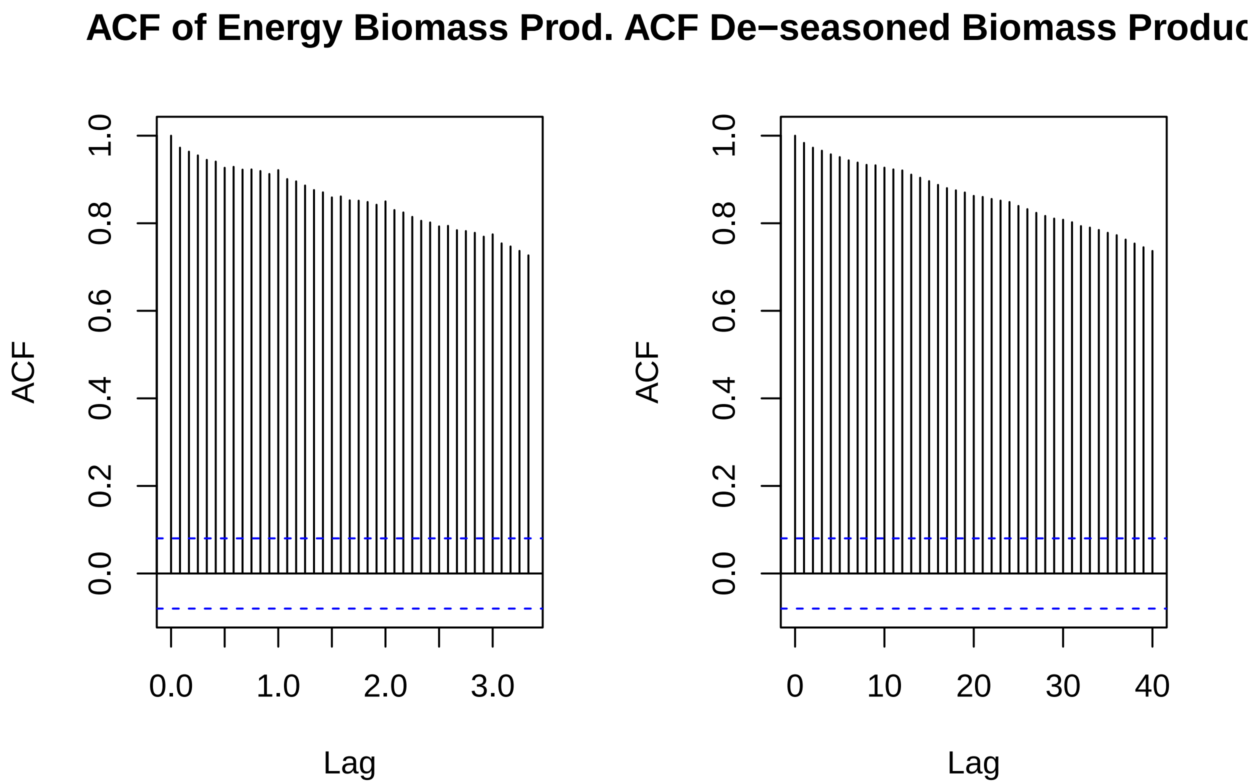



**Q8**

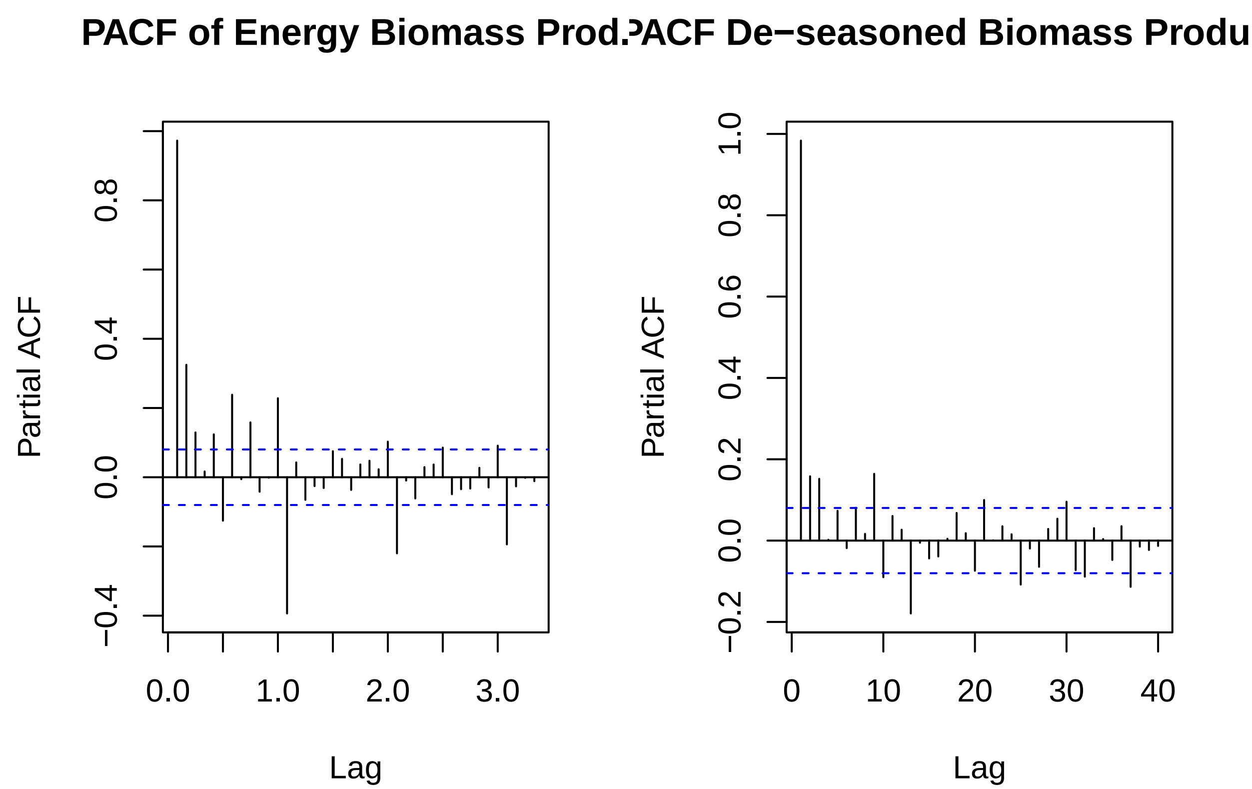
Plot ACF and PACF for the deseason series and compare with the plots from Q1. Did the plots change? How?

> Answer: The Hydroelectric Consumption graphs changed the most dramatically compared to the Q1 plots since significant seasonality was removed. For instance, the ACF flucuating pattern was removed.

```r
par(mfrow = c(1,2))
#Total Biomass Production
acf(ts_energy_data[,1],lag.max=40, type="correlation",plot=TRUE, main = "ACF of Energy Biomass Prod.")
acf(deseason_biomass,lag.max=40, plot=TRUE, main = "ACF De-seasoned Biomass Production")
```

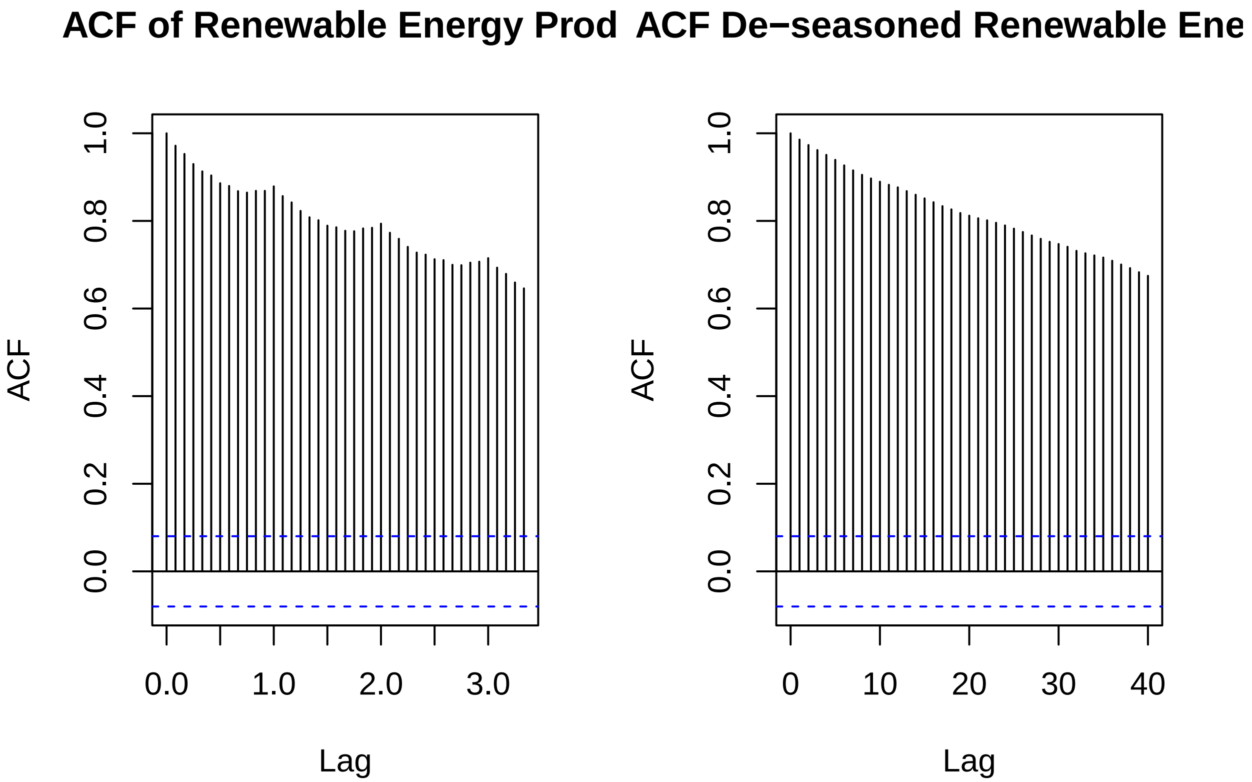



```
pacf(ts_energy_data[,1],lag.max=40, plot=TRUE,  main = "PACF of Energy Biomass Prod.")
pacf(deseason_biomass,lag.max=40, plot=TRUE, main = "PACF De-seasoned Biomass Production")
```

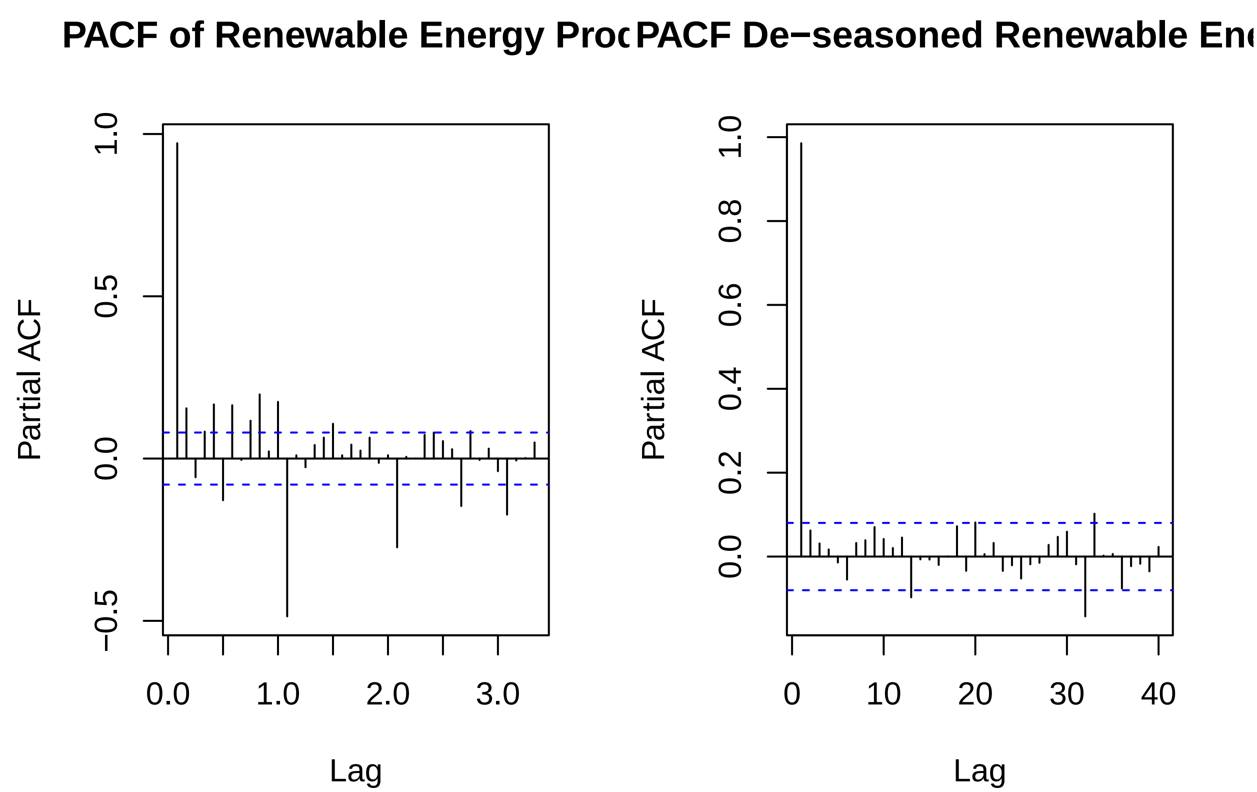



```
#Total Renewable Energy Production
acf(ts_energy_data[,2],lag.max=40, type="correlation",plot=TRUE, main = "ACF of Renewable Energy Prod.")
```

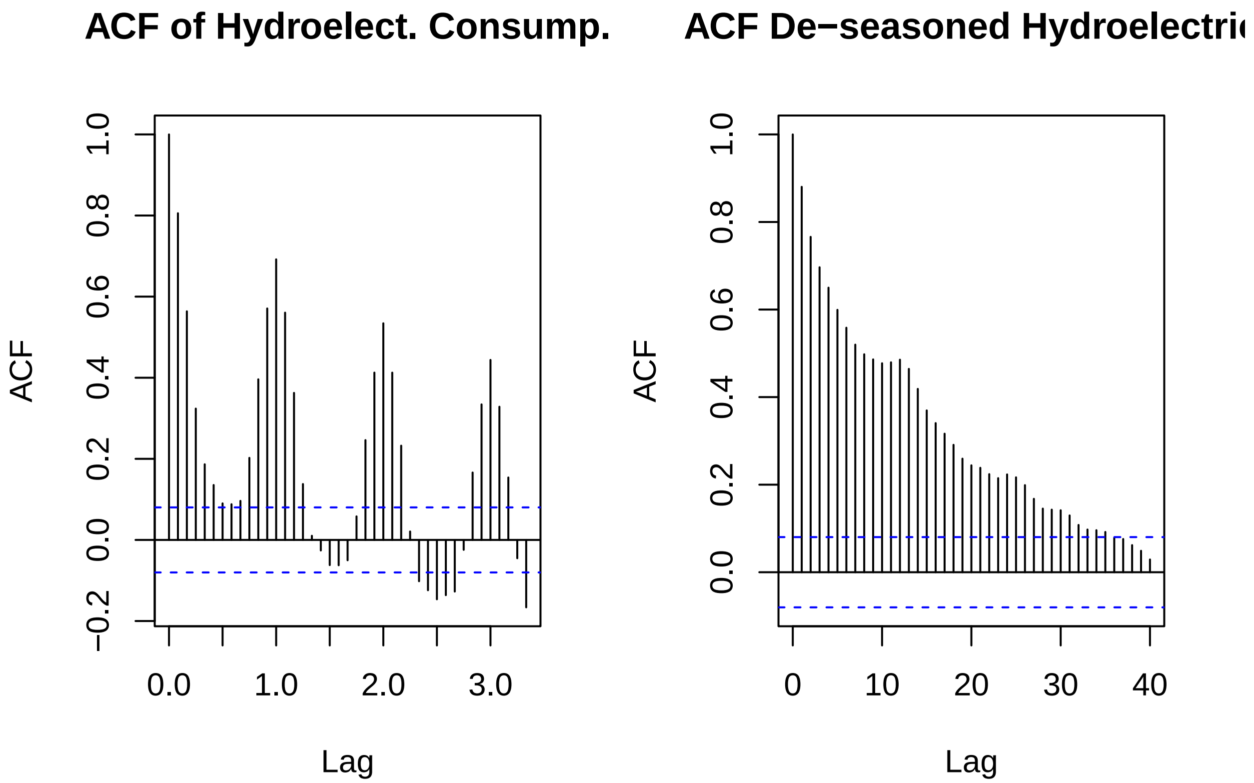
```
acf(deseason_renew,lag.max=40, plot=TRUE, main = "ACF De-seasoned Renewable Energy")
```



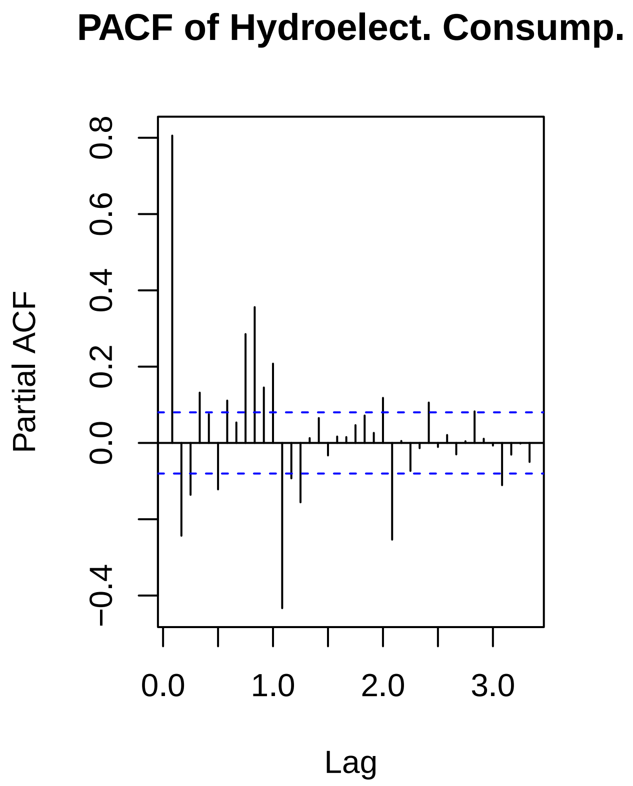

```
pacf(ts_energy_data[,2],lag.max=40, plot=TRUE,  main = "PACF of Renewable Energy Prod.")
pacf(deseason_renew,lag.max=40, plot=TRUE, main = "PACF De-seasoned Renewable Energy")
```

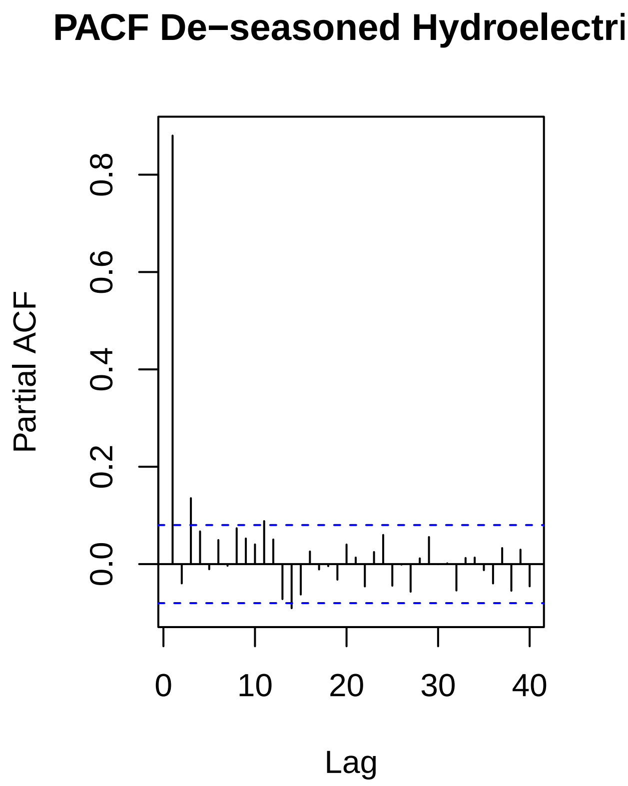


```r
#Hydroelectric Consumption
acf(ts_energy_data[,3],lag.max=40, type="correlation",plot=TRUE, main = "ACF of Hydroelect. Consump.")
acf(deseason_hydro,lag.max=40, plot=TRUE, main = "ACF De-seasoned Hydroelectric")
```




```r
pacf(ts_energy_data[,3],lag.max=40, plot=TRUE,  main = "PACF of Hydroelect. Consump.")
pacf(deseason_hydro,lag.max=40, plot=TRUE, main = "PACF De-seasoned Hydroelectric")
```


PACF of Hydroelect. Consump.
Partial ACF
−0.4
0.0
0.2
0.4
0.6
0.8
0.0
1.0
2.0
3.0
Lag


PACF De−seasoned Hydroelectri
Partial ACF
0.0
0.2
0.4
0.6
0.8
0
10
20
30
40
Lag