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

Assignment 3 - Due date 02/10/23

John Rooney

## 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_Consumpti
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.

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(forecast)
library(tseries)
library(Kendall)
library(grid)
library(gridExtra)
library(cowplot)
library(tidyverse)
```

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

```r
#create data frame
energy_data_df<-energy_data %>%
  select(Month, Total.Biomass.Energy.Production, Total.Renewable.Energy.Production, Hydroelectric.Power
head(energy_data_df)
```

```
##           Month Total.Biomass.Energy.Production
## 1  1973 January                         129.787
## 2 1973 February                         117.338
## 3    1973 March                         129.938
## 4    1973 April                         125.636
## 5      1973 May                         129.834
## 6     1973 June                         125.611
##   Total.Renewable.Energy.Production Hydroelectric.Power.Consumption
## 1                           403.981                         272.703
## 2                           360.900                         242.199
## 3                           400.161                         268.810
## 4                           380.470                         253.185
## 5                           392.141                         260.770
## 6                           377.232                         249.859
```

##Trend Component

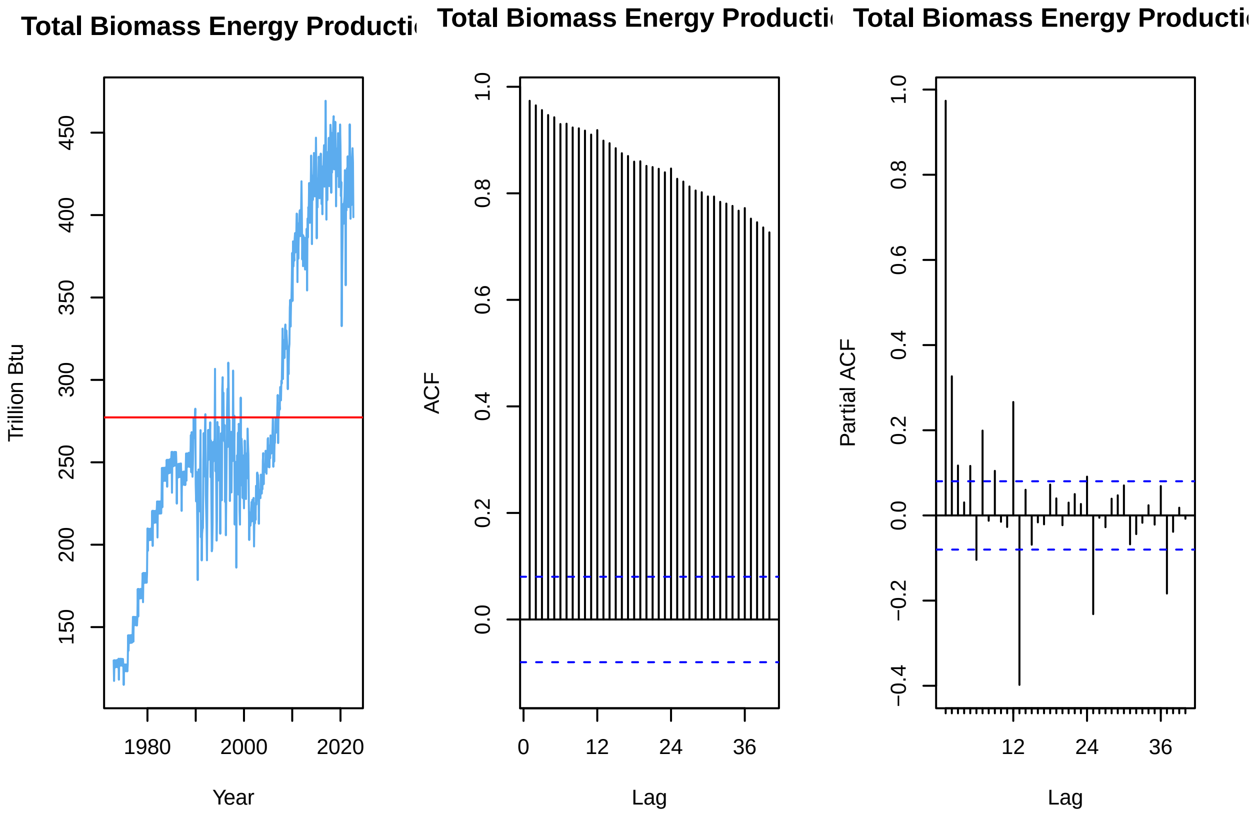
### Q1

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 same code from A2, but I want all three plots on the same window this time. (Hint: use par() function)_

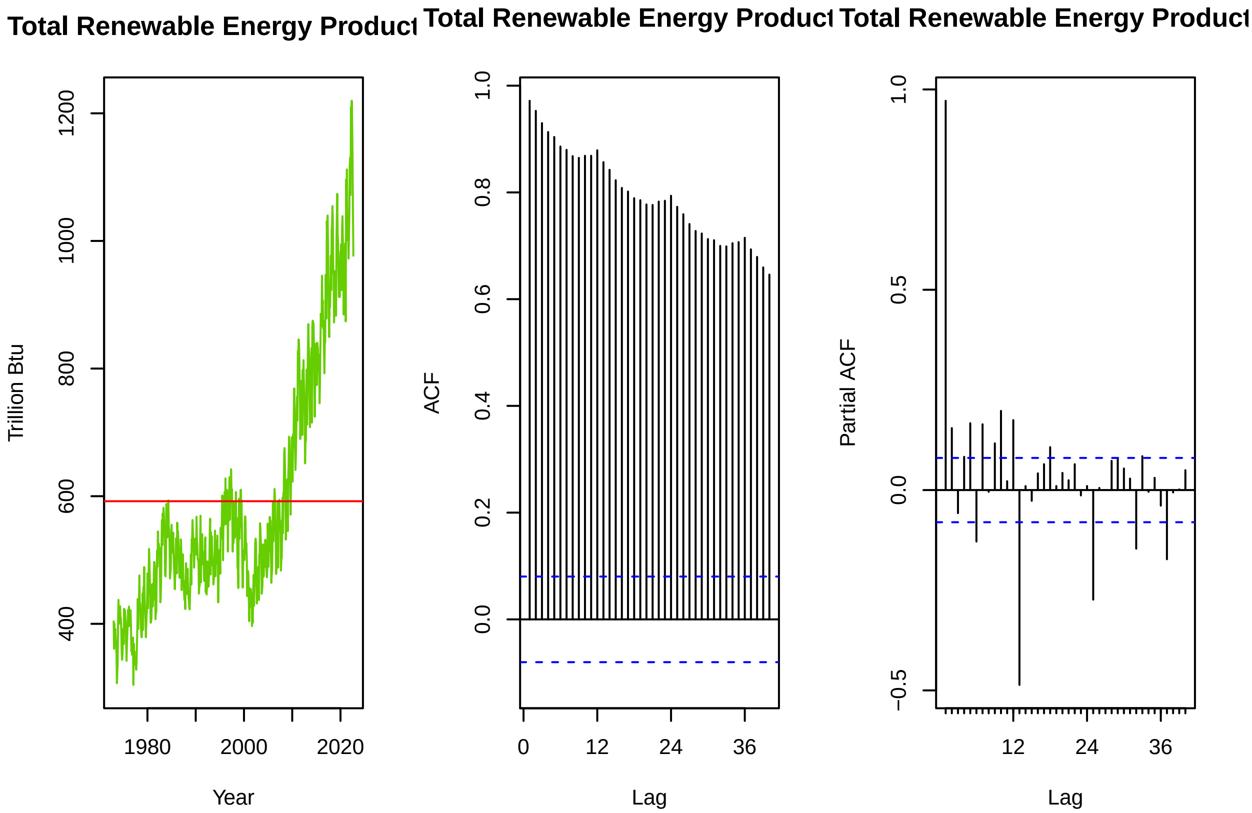
```r
#Plot TS, ACF, PACF
ts_energy_data<-ts(energy_data_df[,2:4], frequency=12,start=c(1973,1))
head(ts_energy_data)
```

```
##          Total.Biomass.Energy.Production Total.Renewable.Energy.Production
## Jan 1973                         129.787                           403.981
## Feb 1973                         117.338                           360.900
## Mar 1973                         129.938                           400.161
## Apr 1973                         125.636                           380.470
## May 1973                         129.834                           392.141
## Jun 1973                         125.611                           377.232
##          Hydroelectric.Power.Consumption
## Jan 1973                         272.703
## Feb 1973                         242.199
## Mar 1973                         268.810
## Apr 1973                         253.185
## May 1973                         260.770
## Jun 1973                         249.859
```

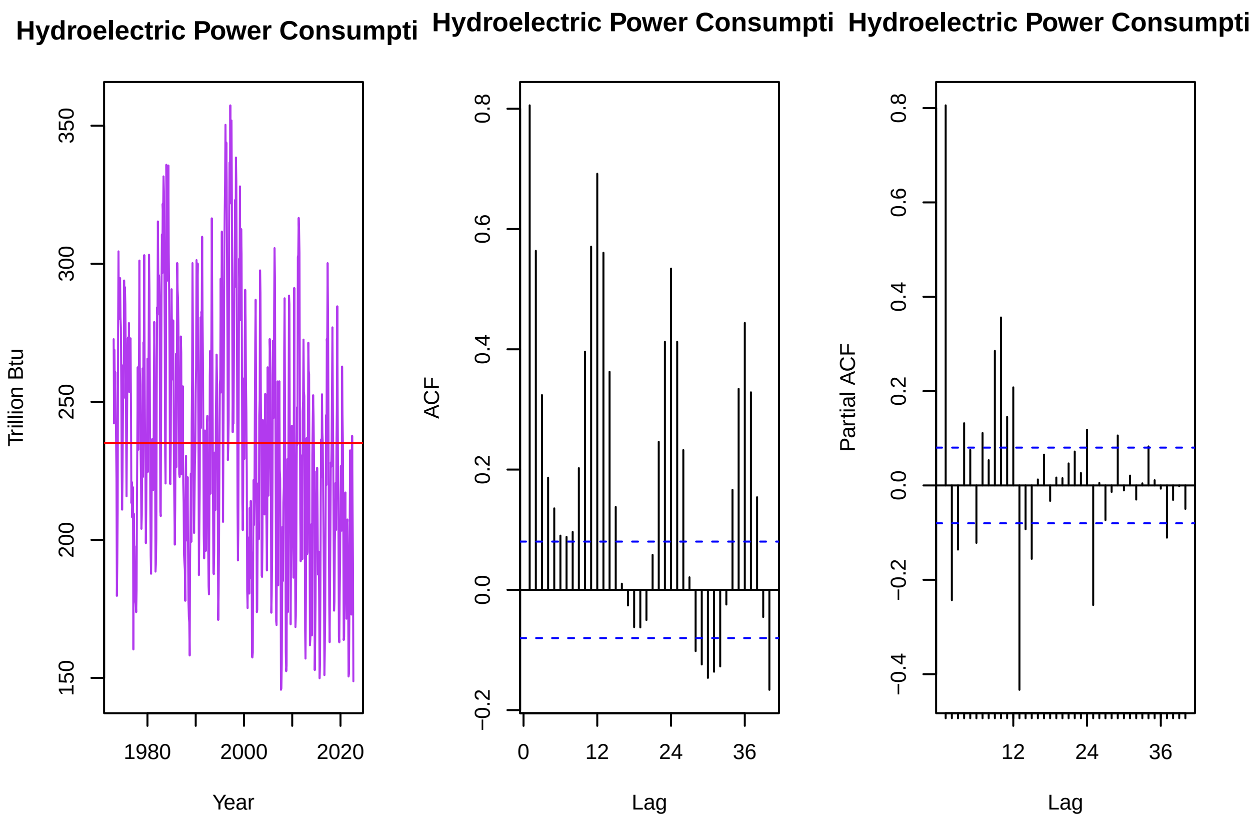
```r
#plot biomass
par(mfrow=c(1,3))
plot(ts_energy_data[,"Total.Biomass.Energy.Production"],type="l",col="steelblue2",xlab="Year",ylab="Tri
abline(h=mean(ts_energy_data[,"Total.Biomass.Energy.Production"]),col="red")
Acf(ts_energy_data[,1],lag.max=40,main=paste("Total Biomass Energy Production",sep=""))
Pacf(ts_energy_data[,1],lag.max=40,main=paste("Total Biomass Energy Production",sep=""))
```




```r
#plot renewable
par(mfrow=c(1,3))
plot(ts_energy_data[,"Total.Renewable.Energy.Production"],type="l",col="chartreuse3",xlab="Year",ylab="
abline(h=mean(ts_energy_data[,"Total.Renewable.Energy.Production"]), col="red")
Acf(ts_energy_data[,2],lag.max=40,main=paste("Total Renewable Energy Production",sep=""))
Pacf(ts_energy_data[,2],lag.max=40,main=paste("Total Renewable Energy Production",sep=""))
```




```
#plot hydro
par(mfrow=c(1,3))
plot(ts_energy_data[,"Hydroelectric.Power.Consumption"],type="l",col="darkorchid2",xlab="Year",ylab="Tri
abline(h=mean(ts_energy_data[,"Hydroelectric.Power.Consumption"]), col="red")
Acf(ts_energy_data[,3],lag.max=40,main=paste("Hydroelectric Power Consumption",sep=""))
Pacf(ts_energy_data[,3],lag.max=40,main=paste("Hydroelectric Power Consumption",sep=""))
```




**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?

The series for Total Biomass Energy Production and Total Renewable Energy Production seem to have a clear positive trend as time goes on. In the case of Total Biomass Energy Production, this is most pronounced from approximately 1975 to 1990 and shortly after 2000 to approximately 2015. In the case of Total Renewable Energy Production, this is most pronounced from 1973 to approximately 1985 and from shortly after 2000 to the present. If a trend exists for Hydroelectric Power Consumption, it is far less clear. Overall, there seems to be an overall negative trend over the length of the time series.

**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.

```
#lm biomass
t<-c(1:597)
biomass_linear_trend_model<-lm(ts_energy_data[,"Total.Biomass.Energy.Production"]~t)
summary(biomass_linear_trend_model)
```

```
##
## Call:
## lm(formula = ts_energy_data[, "Total.Biomass.Energy.Production"] ~
```

```
##     t)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -102.800  -23.994    5.667   32.265   82.192
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.337e+02  3.245e+00   41.22   <2e-16 ***
## t           4.800e-01  9.402e-03   51.05   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 39.59 on 595 degrees of freedom
## Multiple R-squared:  0.8142, Adjusted R-squared:  0.8138
## F-statistic:  2607 on 1 and 595 DF,  p-value: < 2.2e-16
```

```
biomass_beta0=as.numeric(biomass_linear_trend_model$coefficients[1])
biomass_beta1=as.numeric(biomass_linear_trend_model$coefficients[2])

renewable_linear_trend_model<-lm(ts_energy_data[,"Total.Renewable.Energy.Production"]~t)
summary(renewable_linear_trend_model)
```

```
##
## Call:
## lm(formula = ts_energy_data[, "Total.Renewable.Energy.Production"] ~
##     t)
##
## 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 ***
## t             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
```

```
renewable_beta0=as.numeric(renewable_linear_trend_model$coefficients[1])
renewable_beta1=as.numeric(renewable_linear_trend_model$coefficients[2])

hydro_linear_trend_model<-lm(ts_energy_data[,"Hydroelectric.Power.Consumption"]~t)
summary(hydro_linear_trend_model)
```

```
##
## Call:
## lm(formula = ts_energy_data[, "Hydroelectric.Power.Consumption"] ~
##     t)
```

```
##
## 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 ***
## t            -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
```

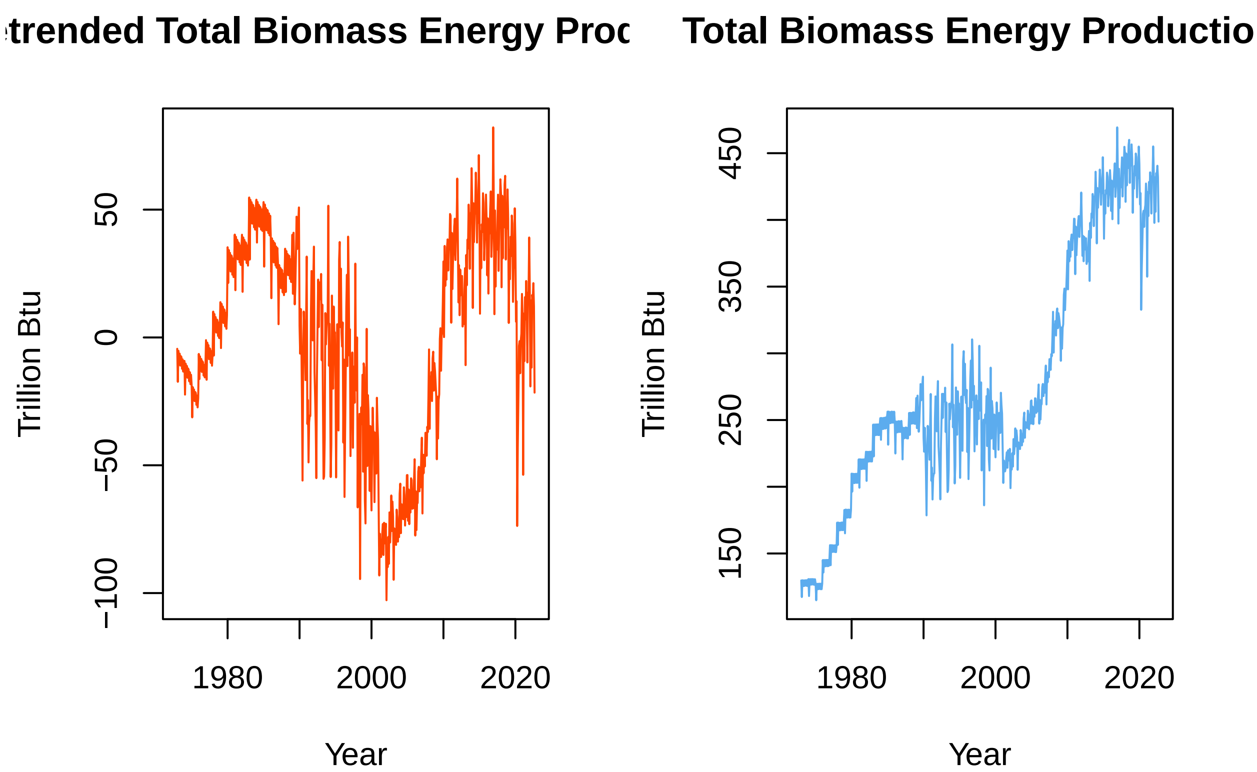
```
hydro_beta0=as.numeric(hydro_linear_trend_model$coefficients[1])
hydro_beta1=as.numeric(hydro_linear_trend_model$coefficients[2])
```

For Total Biomass Energy Production, the linear trend model tells us that the intercept of the trend line is at 133.7 Trillion Btu with a positive slope of.48. The linear trend model for Total Renewable Energy Production tells us that the intercept of the trend line is at 312.25 with a positive slope of.94. The linear trend model for Hydroelectric Power Consumptions tells us that the intercept of the trend line is at 259.9 with a negative slope of -.08.

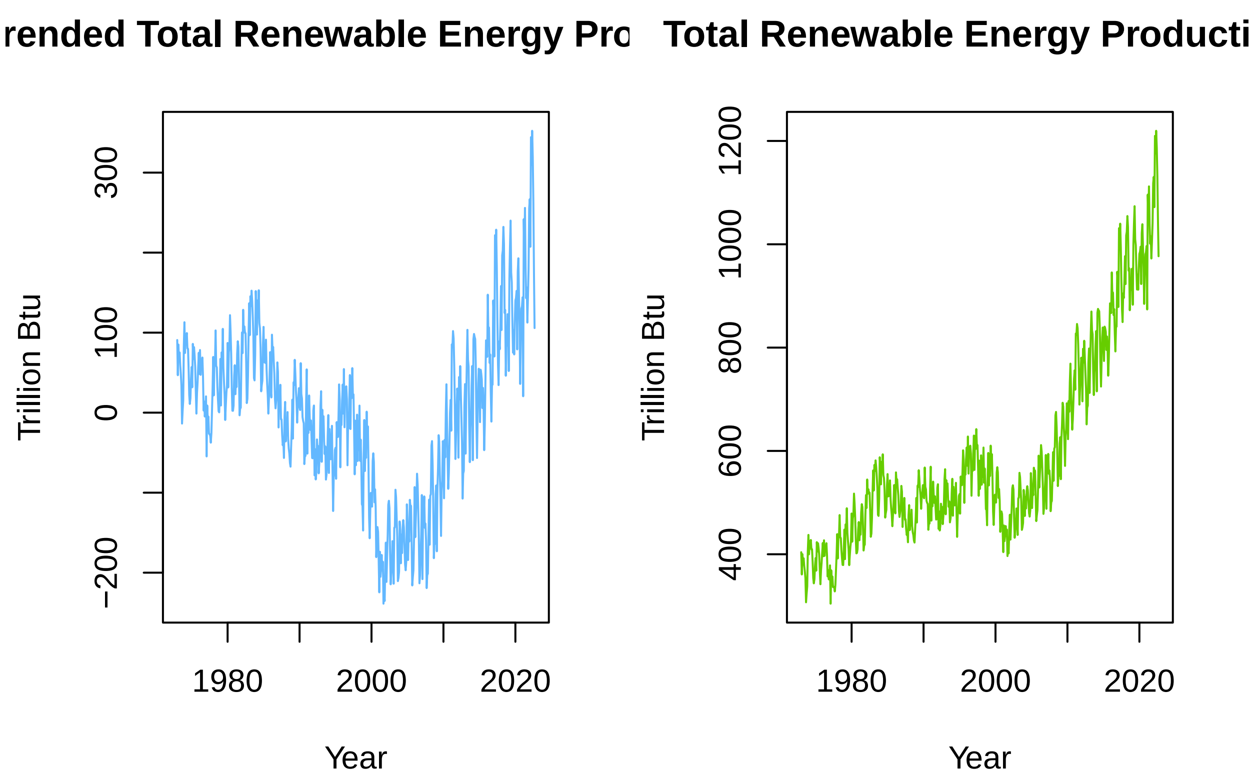
### Q4

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?

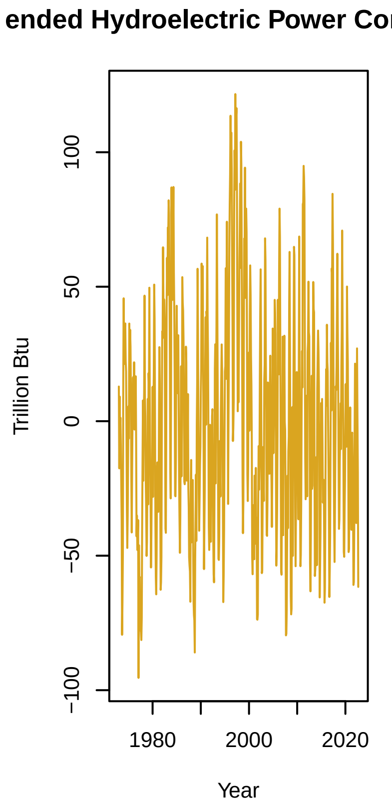
```
#detrend biomass
detrend_biomass <- ts_energy_data[,"Total.Biomass.Energy.Production"]-(biomass_beta0+biomass_beta1*t)

#detrend renewable
detrend_renewable <- ts_energy_data[,"Total.Renewable.Energy.Production"]-(renewable_beta0+renewable_be

#detrend hydro
detrend_hydro <- ts_energy_data[,"Hydroelectric.Power.Consumption"]-(hydro_beta0+hydro_beta1*t)

#plot detrend biomass
par(mfrow=c(1,2))
plot(detrend_biomass,type="l",col="orangered",xlab="Year",ylab="Trillion Btu",main="Detrended Total Bio
plot(ts_energy_data[,"Total.Biomass.Energy.Production"],type="l",col="steelblue2",xlab="Year",ylab="Tri
```

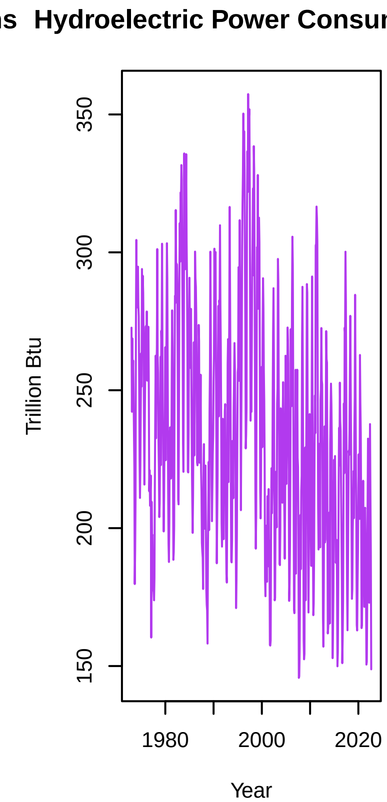



```
#plot detrended renewable
par(mfrow=c(1,2))
plot(detrend_renewable,type="l",col="steelblue1",xlab="Year",ylab="Trillion Btu",main="Detrended Total R
plot(ts_energy_data[,"Total.Renewable.Energy.Production"],type="l",col="chartreuse3",xlab="Year",ylab="T
```



```r
#plot detrended hydro
par(mfrow=c(1,3))
plot(detrend_hydro,type="l",col="goldenrod",xlab="Year",ylab="Trillion Btu",main="Detrended Hydroelectri
plot(ts_energy_data[,"Hydroelectric.Power.Consumption"],type="l",col="darkorchid2",xlab="Year",ylab="Tri
```



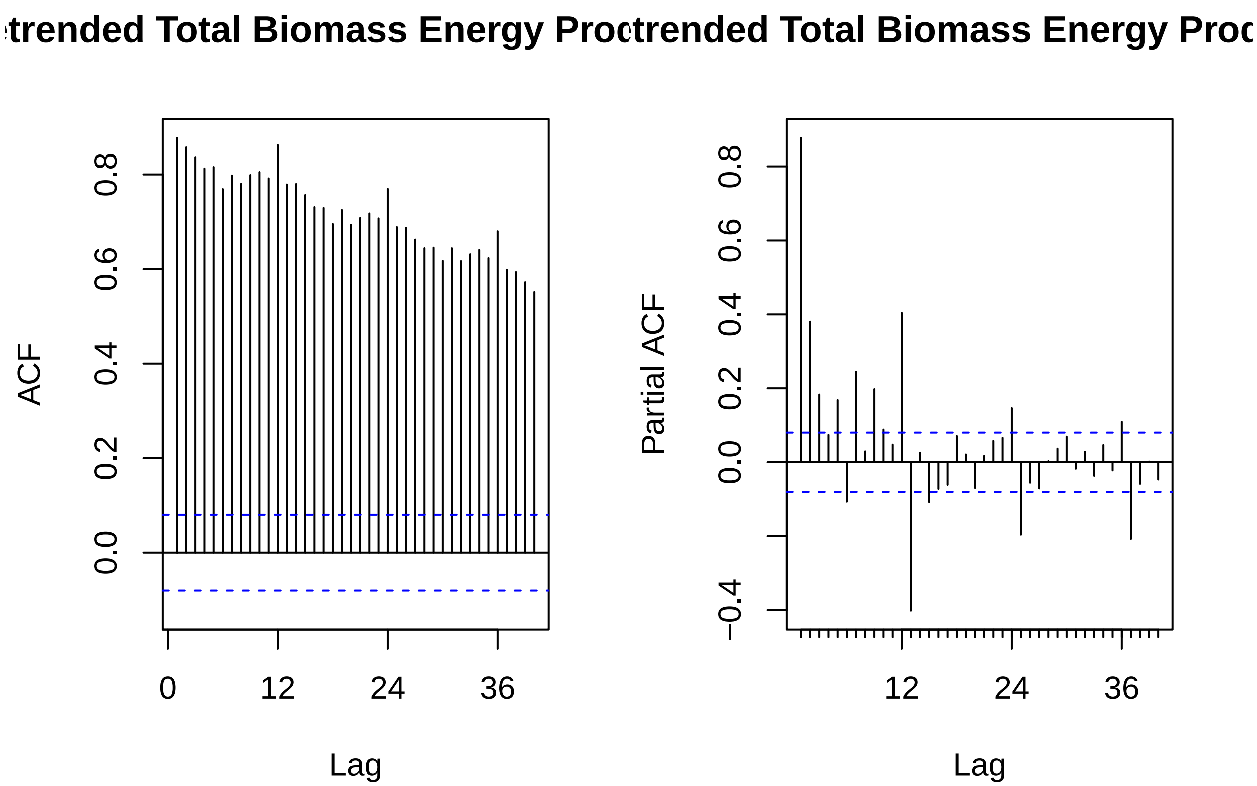




In all three cases, while the overall shapes of the lines are similar the detrended lines show observations with negative values while the original series are all positive values.

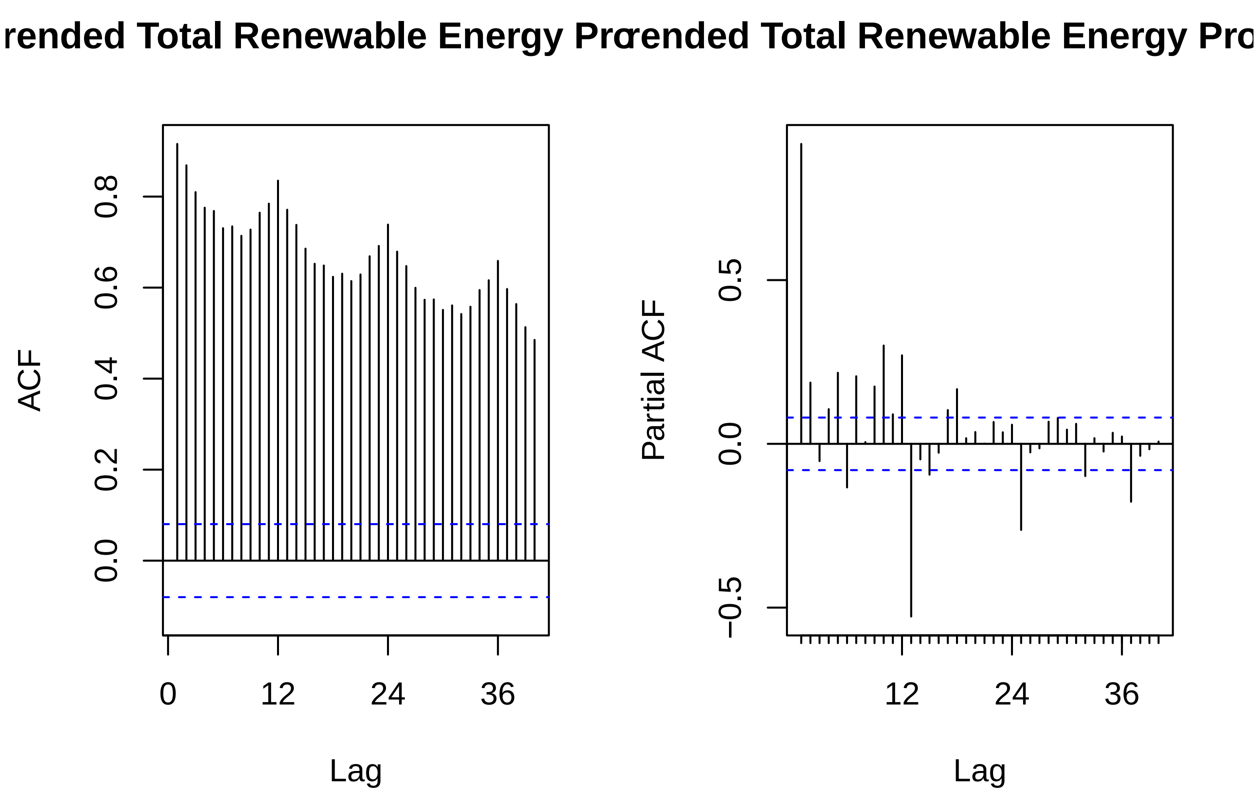
### Q5

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

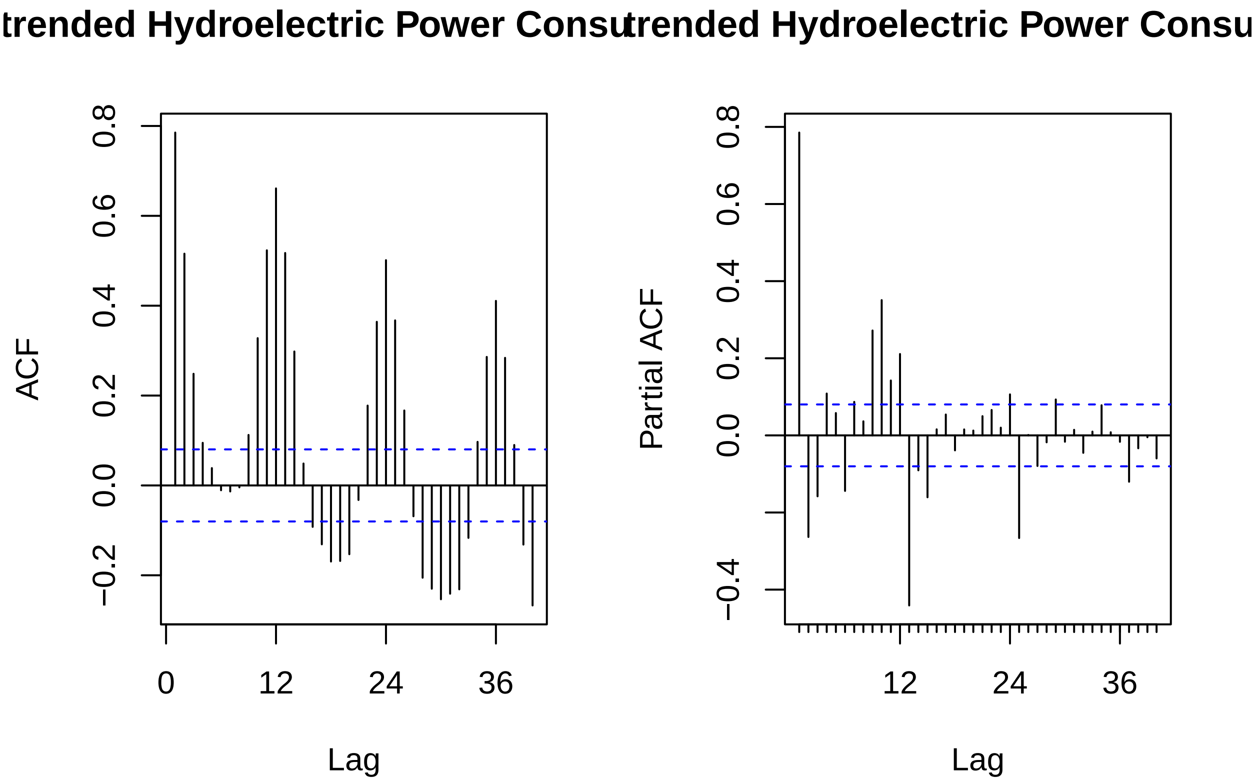
```r
#detrend biomass Acf Pacf
par(mfrow=c(1,2))
Acf(detrend_biomass,lag.max=40,main=paste("Detrended Total Biomass Energy Production",sep=""))
Pacf(detrend_biomass,lag.max=40,main=paste("Detrended Total Biomass Energy Production",sep=""))
```




```
#detrend renewable Acf Pacf
par(mfrow=c(1,2))
Acf(detrend_renewable,lag.max=40,main=paste("Detrended Total Renewable Energy Production",sep=""))
Pacf(detrend_renewable,lag.max=40,main=paste("Detrended Total Renewable Energy Production",sep=""))
```



```
#detrend hydro Acf Pacf
par(mfrow=c(1,2))
Acf(detrend_hydro,lag.max=40,main=paste("Detrended Hydroelectric Power Consumption",sep=""))
Pacf(detrend_hydro,lag.max=40,main=paste("Detrended Hydroelectric Power Consumption",sep=""))
```




For biomass, the detrended Acf and Pacf seem similar although the Acf for the detrended series seems to start at a lower value and the seasonality seems more pronounced. The same is true for renewable energy. For hydro power consumption, more of the negative values in the detrended series are significant.

## 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.

```
#fit seasonal means model for biomass
biomass_dummies <- seasonaldummy(ts_energy_data[,"Total.Biomass.Energy.Production"])
biomass_seas_means_model=lm(ts_energy_data[,"Total.Biomass.Energy.Production"]~biomass_dummies)
summary(biomass_seas_means_model)
```

```
##
## Call:
## lm(formula = ts_energy_data[, "Total.Biomass.Energy.Production"] ~
```

```
##     biomass_dummies)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -160.74  -53.67  -24.36   90.73  181.34
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)          288.020     13.163  21.881   <2e-16 ***
## biomass_dummiesJan    -1.793     18.522  -0.097   0.9229
## biomass_dummiesFeb   -31.102     18.522  -1.679   0.0936 .
## biomass_dummiesMar    -9.104     18.522  -0.492   0.6232
## biomass_dummiesApr   -21.502     18.522  -1.161   0.2462
## biomass_dummiesMay   -14.238     18.522  -0.769   0.4424
## biomass_dummiesJun   -19.602     18.522  -1.058   0.2904
## biomass_dummiesJul    -3.674     18.522  -0.198   0.8428
## biomass_dummiesAug    -0.612     18.522  -0.033   0.9737
## biomass_dummiesSep   -13.335     18.522  -0.720   0.4718
## biomass_dummiesOct    -4.030     18.615  -0.216   0.8287
## biomass_dummiesNov    -9.849     18.615  -0.529   0.5970
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 92.14 on 585 degrees of freedom
## Multiple R-squared:  0.01018,    Adjusted R-squared:  -0.008437
## F-statistic: 0.5467 on 11 and 585 DF,  p-value: 0.8714
```

```
biomass_beta_int=biomass_seas_means_model$coefficients[1]
biomass_beta_coeff=biomass_seas_means_model$coefficients[2:12]

#fit seasonal means model for renewable
renewable_dummies <- seasonaldummy(ts_energy_data[,"Total.Renewable.Energy.Production"])
renewable_seas_means_model=lm(ts_energy_data[,"Total.Renewable.Energy.Production"]~renewable_dummies)
summary(renewable_seas_means_model)
```

```
##
## Call:
## lm(formula = ts_energy_data[, "Total.Renewable.Energy.Production"] ~
##     renewable_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 ***
## renewable_dummiesJan    11.468     38.358   0.299    0.765
## renewable_dummiesFeb   -41.456     38.358  -1.081    0.280
## renewable_dummiesMar    23.130     38.358   0.603    0.547
## renewable_dummiesApr     9.959     38.358   0.260    0.795
## renewable_dummiesMay    38.853     38.358   1.013    0.312
## renewable_dummiesJun    20.378     38.358   0.531    0.595
## renewable_dummiesJul     8.298     38.358   0.216    0.829
```

```
## renewable_dummiesAug  -19.450     38.358  -0.507    0.612
## renewable_dummiesSep  -63.770     38.358  -1.662    0.097 .
## renewable_dummiesOct  -52.612     38.551  -1.365    0.173
## renewable_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
```

```
renewable_beta_int=renewable_seas_means_model$coefficients[1]
renewable_beta_coeff=renewable_seas_means_model$coefficients[2:12]

#fit seasonal means model for hydro
hydro_dummies <- seasonaldummy(ts_energy_data[,"Hydroelectric.Power.Consumption"])
hydro_seas_means_model=lm(ts_energy_data[,"Hydroelectric.Power.Consumption"]~hydro_dummies)
summary(hydro_seas_means_model)
```

```
##
## Call:
## lm(formula = ts_energy_data[, "Hydroelectric.Power.Consumption"] ~
##     hydro_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 ***
## hydro_dummiesJan   13.594      6.864   1.981  0.04811 *
## hydro_dummiesFeb   -8.254      6.864  -1.203  0.22964
## hydro_dummiesMar   19.980      6.864   2.911  0.00374 **
## hydro_dummiesApr   15.649      6.864   2.280  0.02297 *
## hydro_dummiesMay   39.210      6.864   5.713 1.77e-08 ***
## hydro_dummiesJun   31.209      6.864   4.547 6.61e-06 ***
## hydro_dummiesJul   10.436      6.864   1.520  0.12895
## hydro_dummiesAug  -17.909      6.864  -2.609  0.00931 **
## hydro_dummiesSep  -50.173      6.864  -7.310 8.82e-13 ***
## hydro_dummiesOct  -48.262      6.898  -6.996 7.22e-12 ***
## hydro_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
```

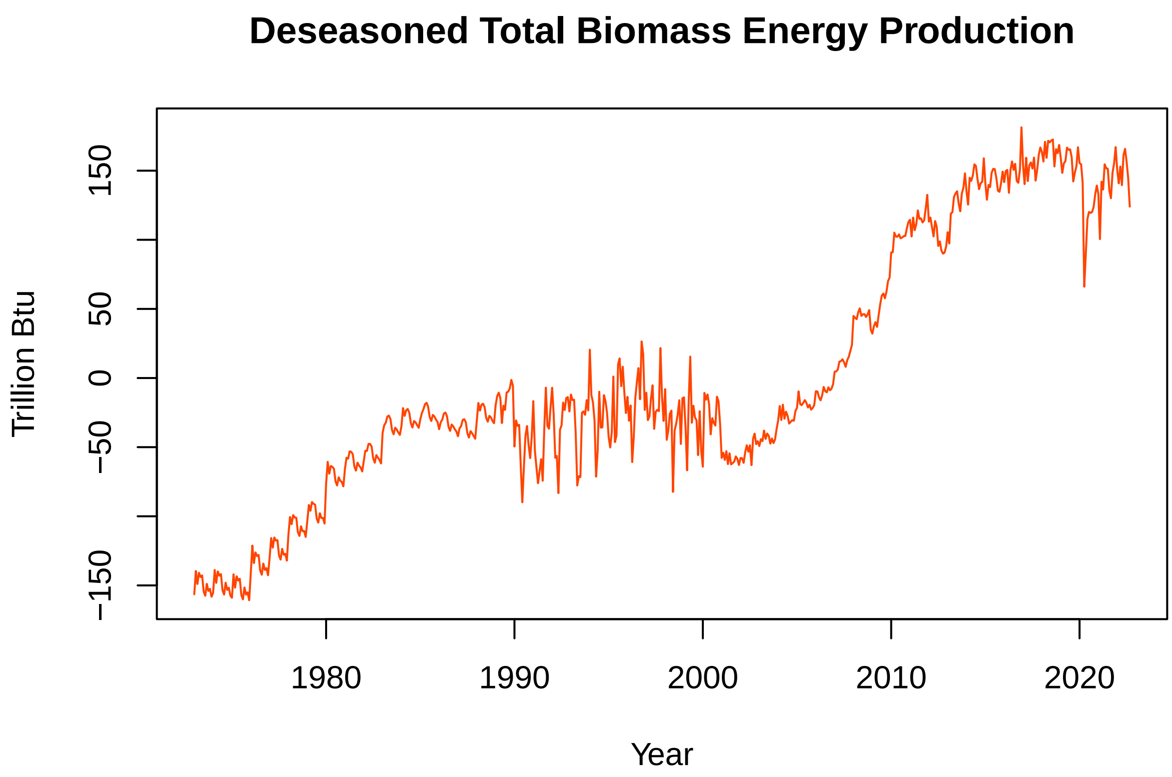
```
hydro_beta_int=hydro_seas_means_model$coefficients[1]
hydro_beta_coeff=hydro_seas_means_model$coefficients[2:12]
```

The biomass energy production seasonal means model shows an intercept value of 288.02 and all negative coefficient values for the dummy variables. The renewable energy production seasonal means model shows an intercept value of 601.02 and negative coefficient values in February and August-November for the dummy variables. The same pattern is true for the hydro power consumption except the intercept value is 237.23.

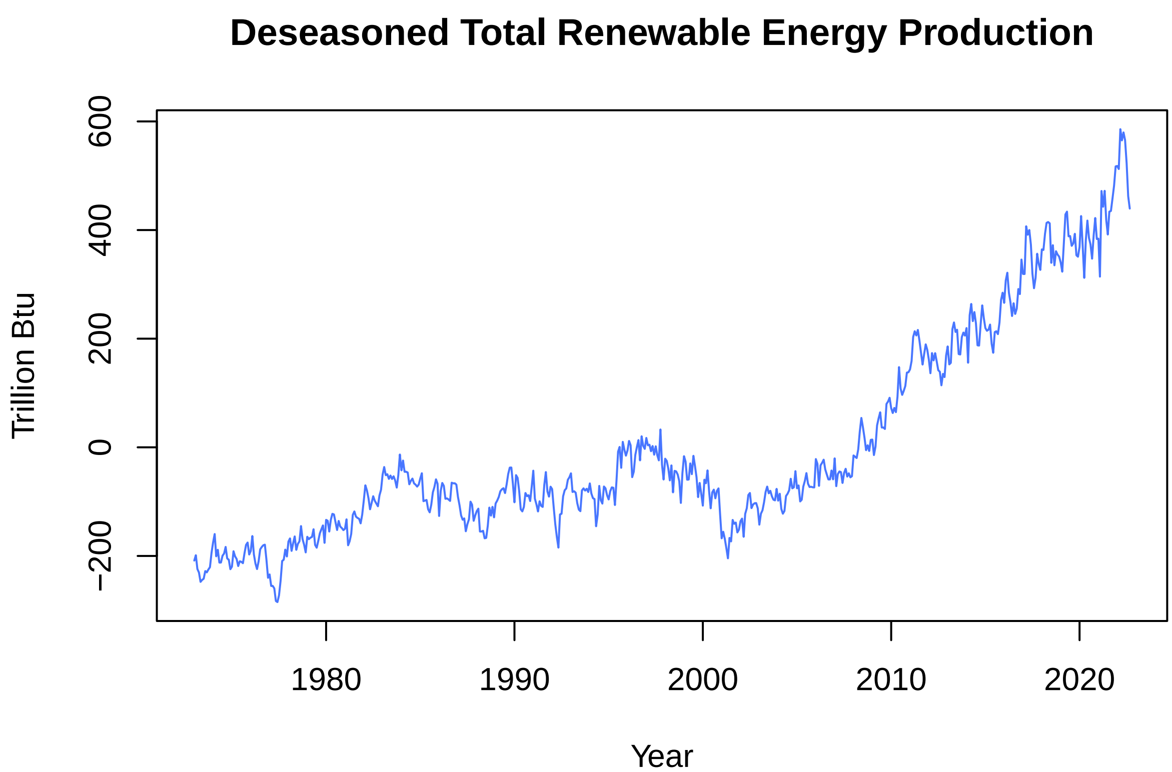
### Q7

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?

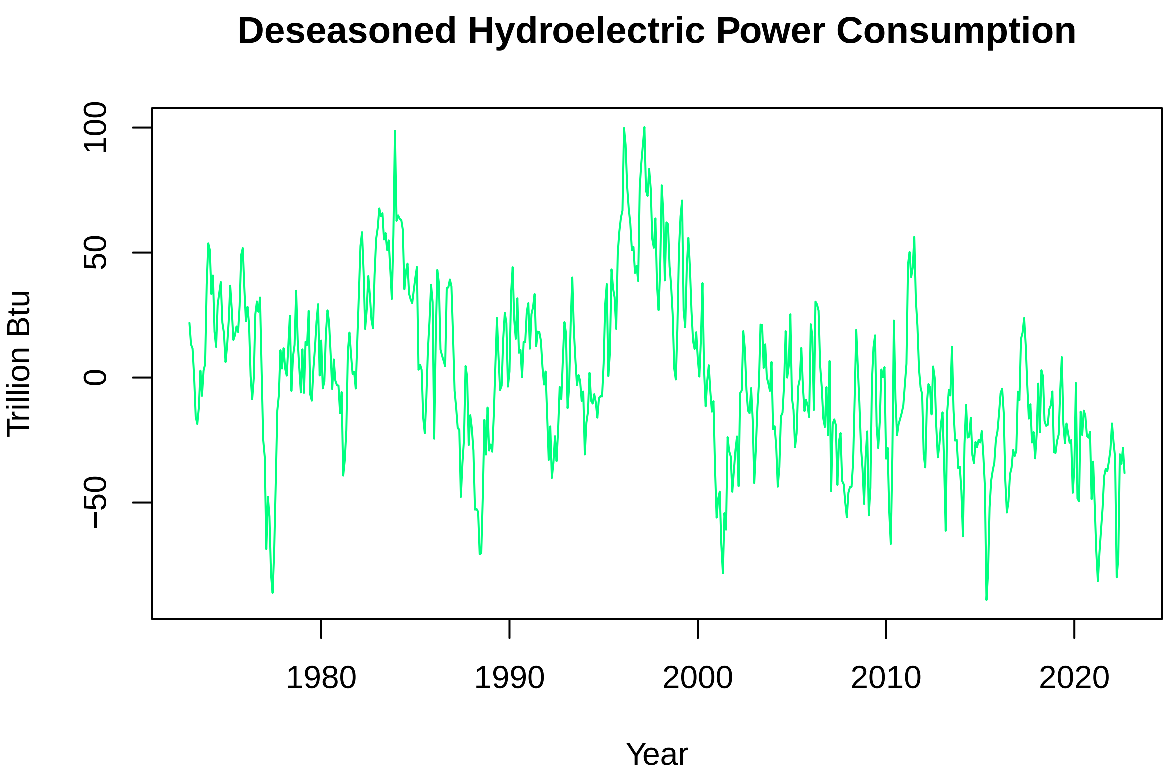
```
#compute seasonal component
biomass_seas_comp=array(0,597)
for(i in 1:597){
  biomass_seas_comp[i]=(biomass_beta_int+biomass_beta_coeff%*%biomass_dummies[i,])
}

renewable_seas_comp=array(0,597)
for(i in 1:597){
  renewable_seas_comp[i]=(renewable_beta_int+renewable_beta_coeff%*%renewable_dummies[i,])
}

hydro_seas_comp=array(0,597)
for(i in 1:597){
  hydro_seas_comp[i]=(hydro_beta_int+hydro_beta_coeff%*%hydro_dummies[i,])
}

#Removing seasonal component
deseason_biomass_energy_data <- ts_energy_data[,"Total.Biomass.Energy.Production"]-biomass_seas_comp

deseason_renewable_energy_data <- ts_energy_data[,"Total.Renewable.Energy.Production"]-renewable_seas_c

deseason_hydro_energy_data <- ts_energy_data[,"Hydroelectric.Power.Consumption"]-hydro_seas_comp

#plot deseason

plot(deseason_biomass_energy_data,type="l",col="orangered",xlab="Year",ylab="Trillion Btu",main="Deseas
```




```
plot(deseason_renewable_energy_data,type="l",col="royalblue1",xlab="Year",ylab="Trillion Btu",main="Dese
```




```
plot(deseason_hydro_energy_data,type="l",col="springgreen",xlab="Year",ylab="Trillion Btu",main="Deseas
```

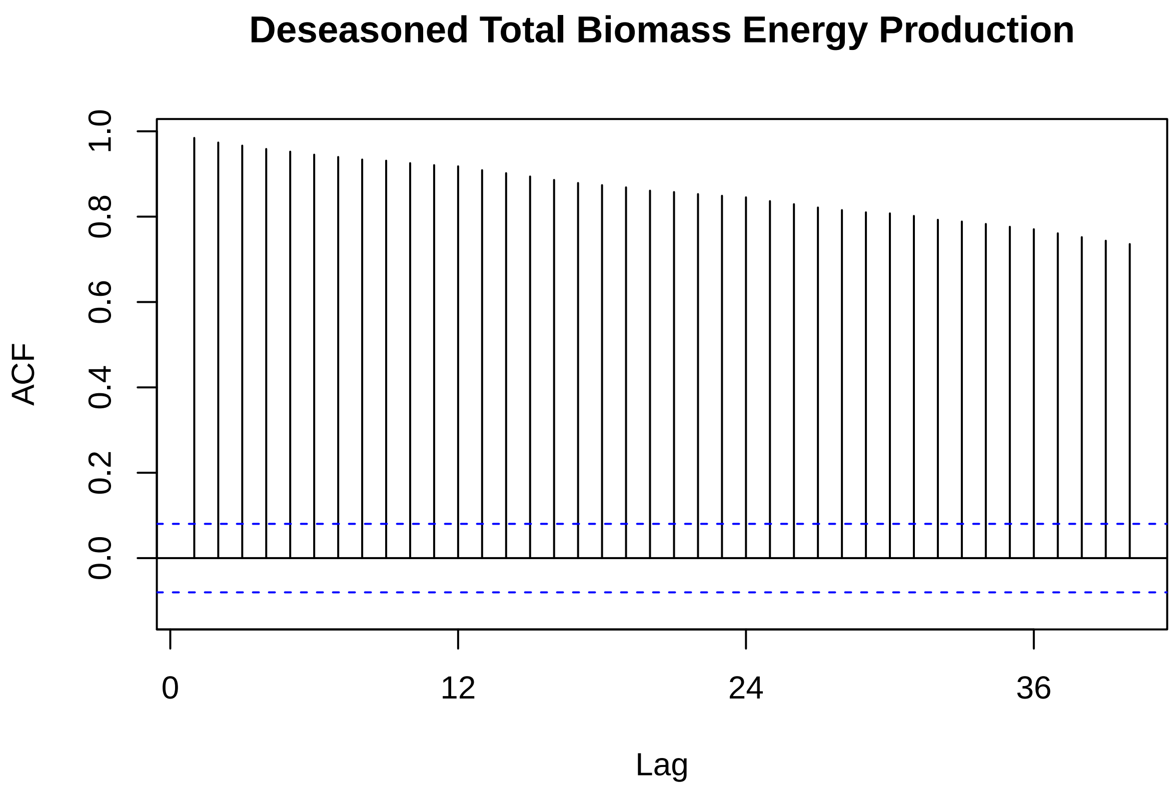



I don't believe anything really changed for the biomass and renewable series. However, the hydro changed significantly, with all values shown being positive and generally between 150 and 350 trillion Btu

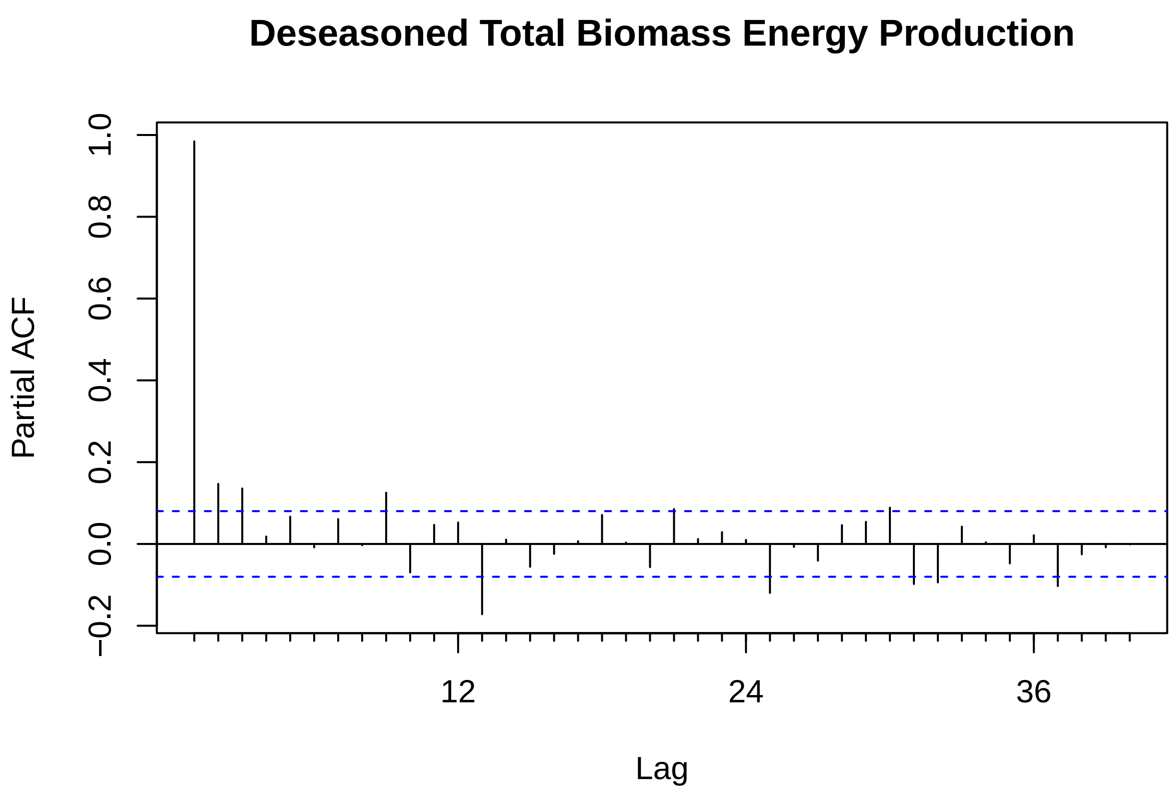
**Q8**

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

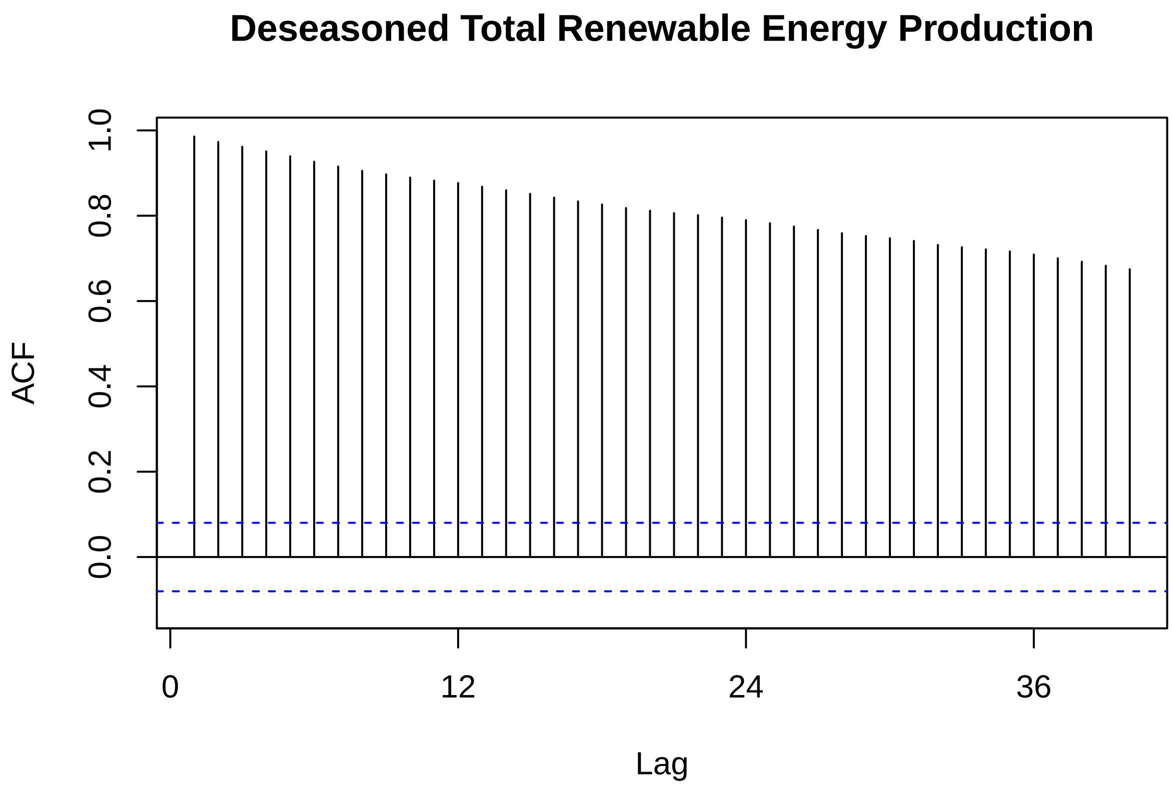
```
#Acf Pacf deseason
Acf(deseason_biomass_energy_data,lag.max=40,main=paste("Deseasoned Total Biomass Energy Production",sep
```



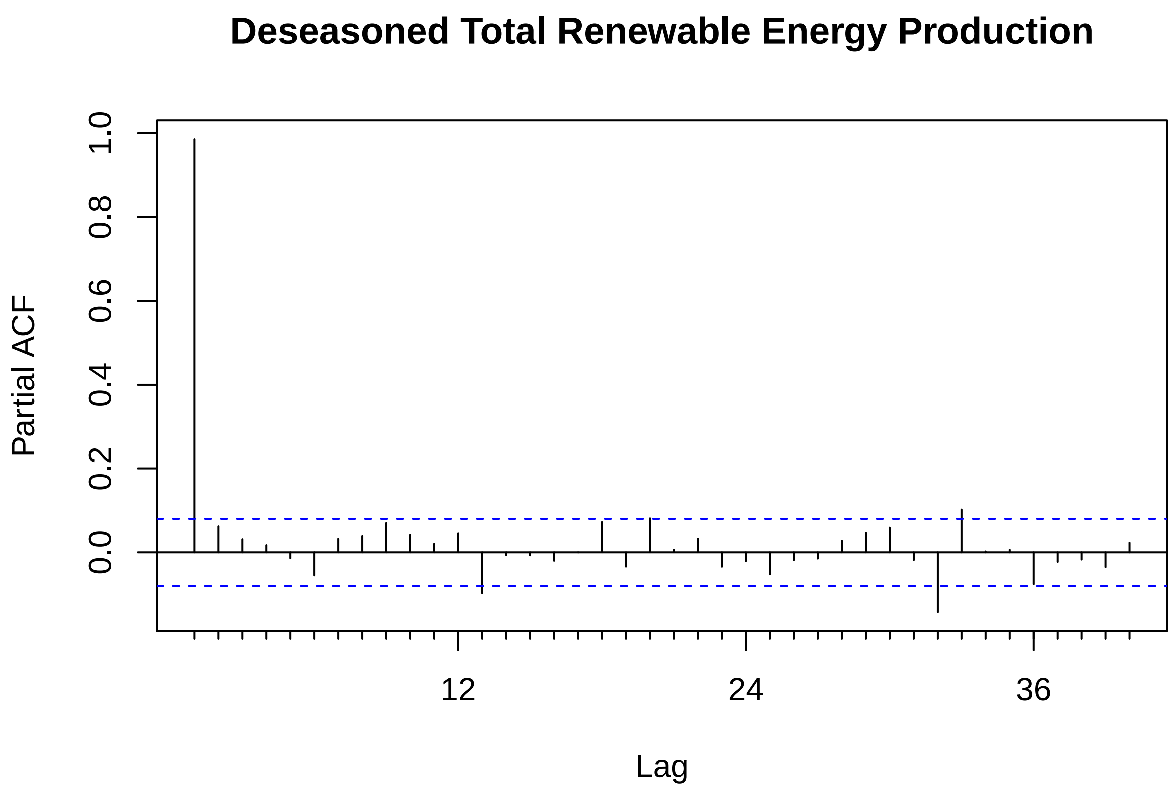

```
Pacf(deseason_biomass_energy_data,lag.max=40,main=paste("Deseasoned Total Biomass Energy Production",se
```



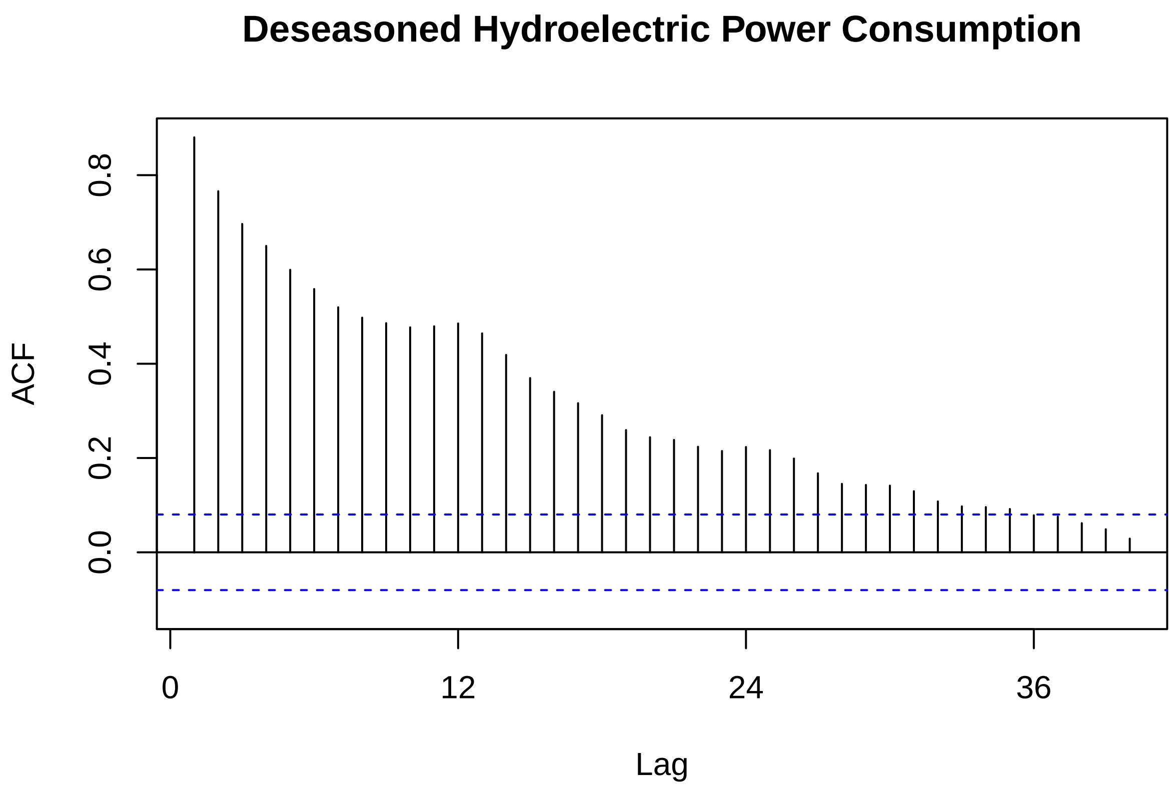
```
Acf(deseason_renewable_energy_data,lag.max=40,main=paste("Deseasoned Total Renewable Energy Production"
```



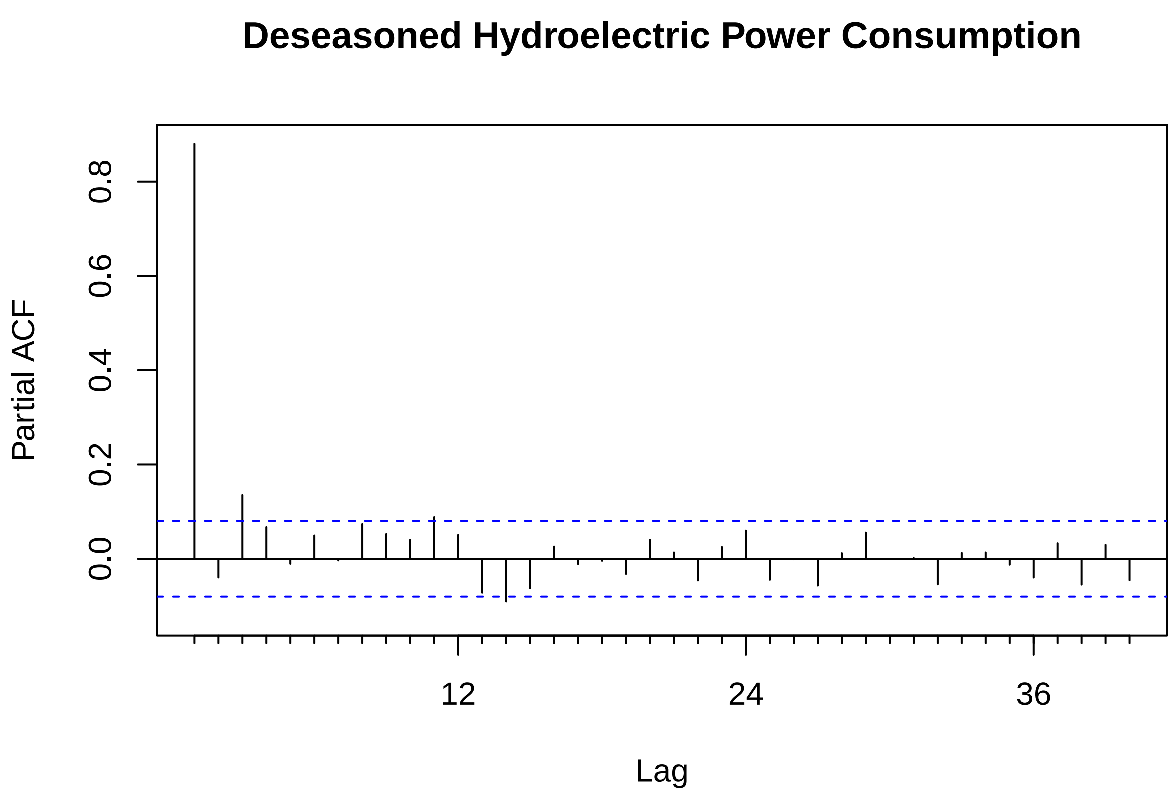

```
Pacf(deseason_renewable_energy_data,lag.max=40,main=paste("Deseasoned Total Renewable Energy Production
```



```
Acf(deseason_hydro_energy_data,lag.max=40,main=paste("Deseasoned Hydroelectric Power Consumption",sep="
```




```
Pacf(deseason_hydro_energy_data,lag.max=40,main=paste("Deseasoned Hydroelectric Power Consumption",sep=
```



For biomass and renewable energy production, the Acf plots seemed to change to show a general decline without the seasonal increase seen in the original data. Pacf plots did not change signficantly, although fewer values were shown to be significant. I'm not sure I'm seeing much of a change for hydro.