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

Assignment 2 - Due date 02/03/23

Kelsey Husted

Submission Instructions

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_Sp22.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).

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

R packages

R packages needed for this assignment:"forecast","tseries", and "dplyr". 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(dplyr)
library(tseries)
library(forecast)
library(lubridate)
library(ggplot2)
```

Data set information

Consider the data provided in the spreadsheet "Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source.x on our **Data** folder. The data comes from the US Energy Information and Administration and corresponds to the December 2022 Monthly Energy Review. The spreadsheet is ready to be used. You will also find a *.csv* version of the data "Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source-Edit.csv". You may use the function *read.table*() to import the *.csv* data in R. Or refer to the file "M2_ImportingData_CSV_XLSX.Rmd" in our Lessons folder for functions that are better suited for importing the *.xlsx*.

```
#Check wd
getwd()
```

```
## [1] "/Users/kelseyhusted 1/Desktop/Time Series Analysis/ENV790TSA"
```

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

Question 1

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. Use the command head() to verify your data.

```
#select specified columns; will include the time column to practice reformatting date
energy_data <- energy_data %>%
  select(1,4:6)
head(energy_data)
```

```
##            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
```

```
#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.Production 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
```

Question 2

Transform your data frame in a time series object and specify the starting point and frequency of the time series using the function ts().

```
#ts transformation
ts_energy_data <- ts(energy_data[,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
```

Question 3

Compute mean and standard deviation for these three series.

```
#Total Biomass Energy Production
Biomass_mean <- mean(ts_energy_data[,1])
Biomass_sd <- sd(ts_energy_data[,1])
Biomass_mean
```

```
## [1] 277.2525
```

```
Biomass_sd
```

```
## [1] 91.75367
```

```
#Total Renewable Energy Production
Renewable_mean <- mean(ts_energy_data[,2])
Renewable_sd <- sd(ts_energy_data[,2])
Renewable_mean
```

```
## [1] 592.1583
```

```
Renewable_sd
```

```
## [1] 191.7978
```

```
#Hydroelectric Power Consumption
Hydroelect_mean <- mean(ts_energy_data[,3])
Hydroelect_sd <- sd(ts_energy_data[,3])
Hydroelect_mean
```

```
## [1] 235.1146
```

```
Hydroelect_sd
```

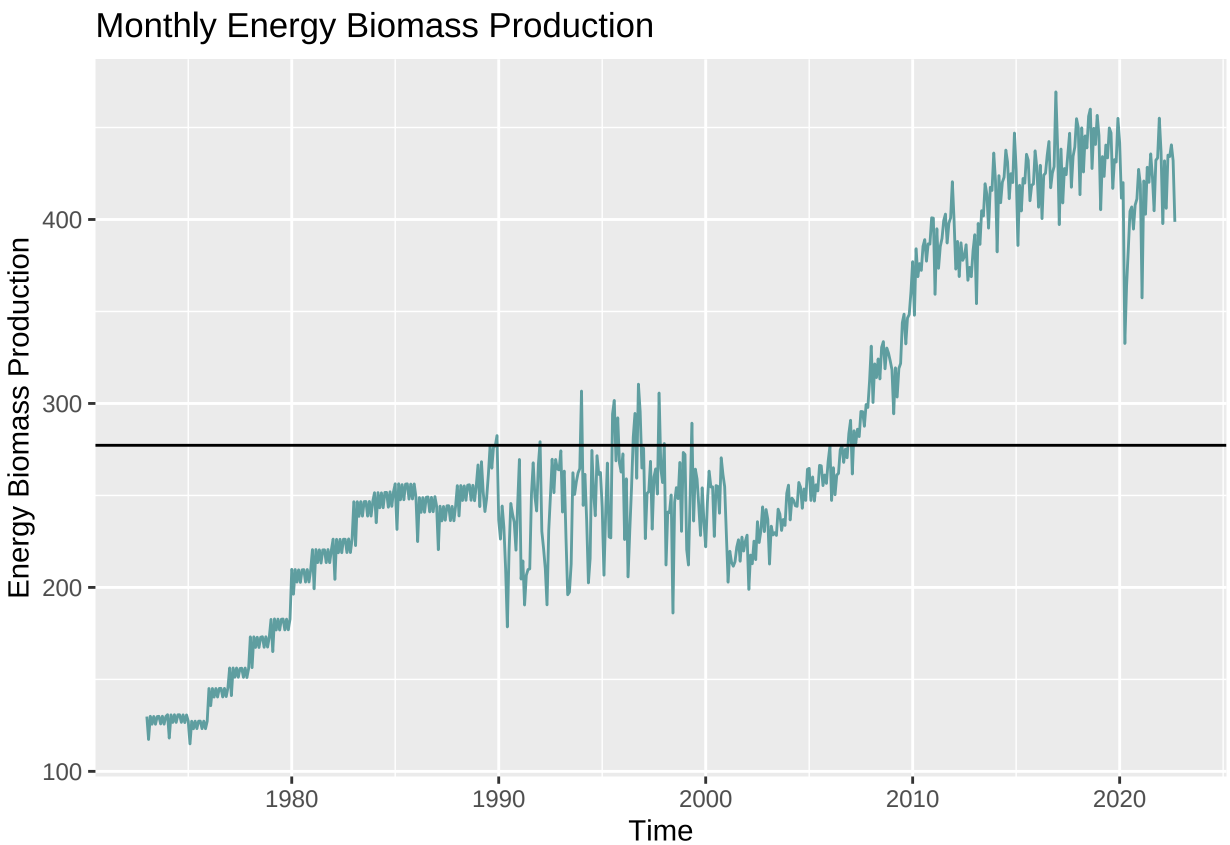
```
## [1] 44.16116
```

Question 4

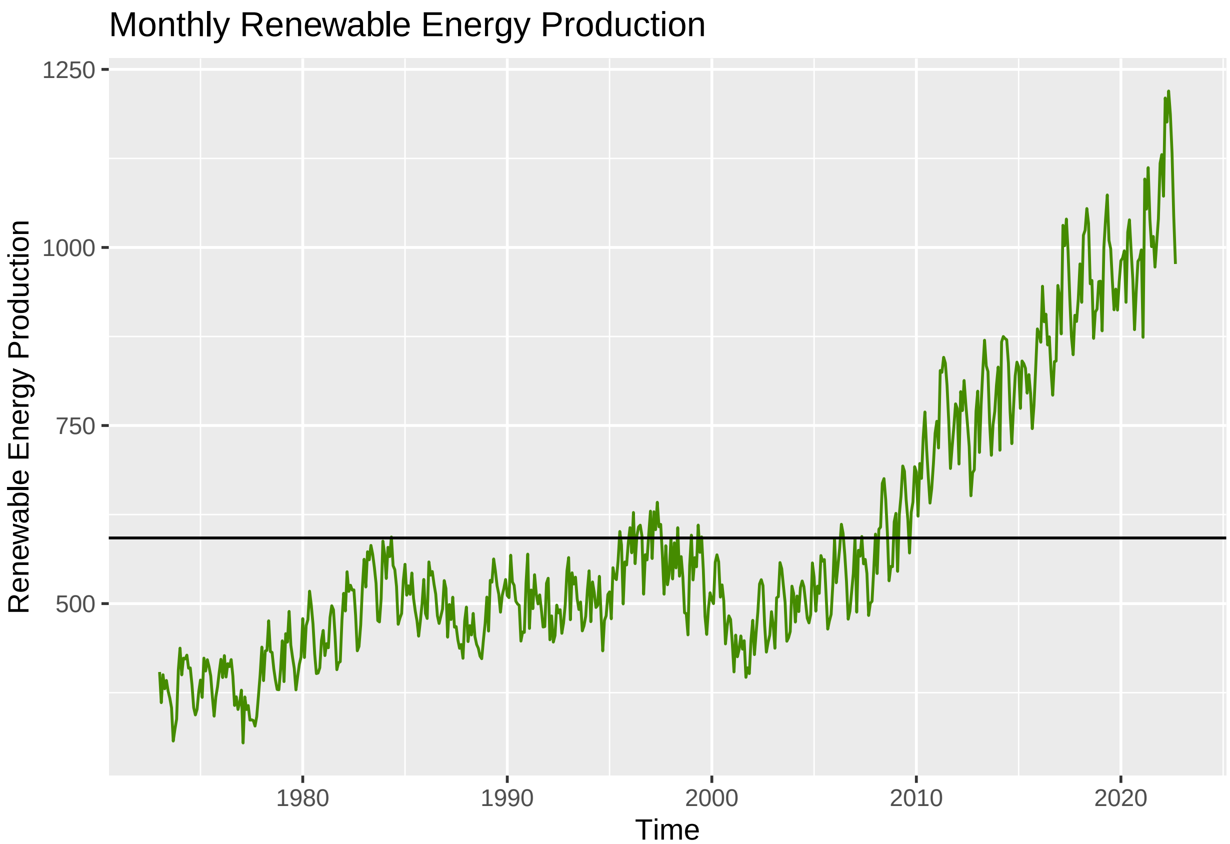
Display and interpret the time series plot for each of these variables. Try to make your plot as informative as possible by writing titles, labels, etc. For each plot add a horizontal line at the mean of each series in a different color.

```
#Total Biomass Energy Production
autoplot(ts_energy_data[,1],main = "Monthly Energy Biomass Production",
         col = "cadetblue",
```

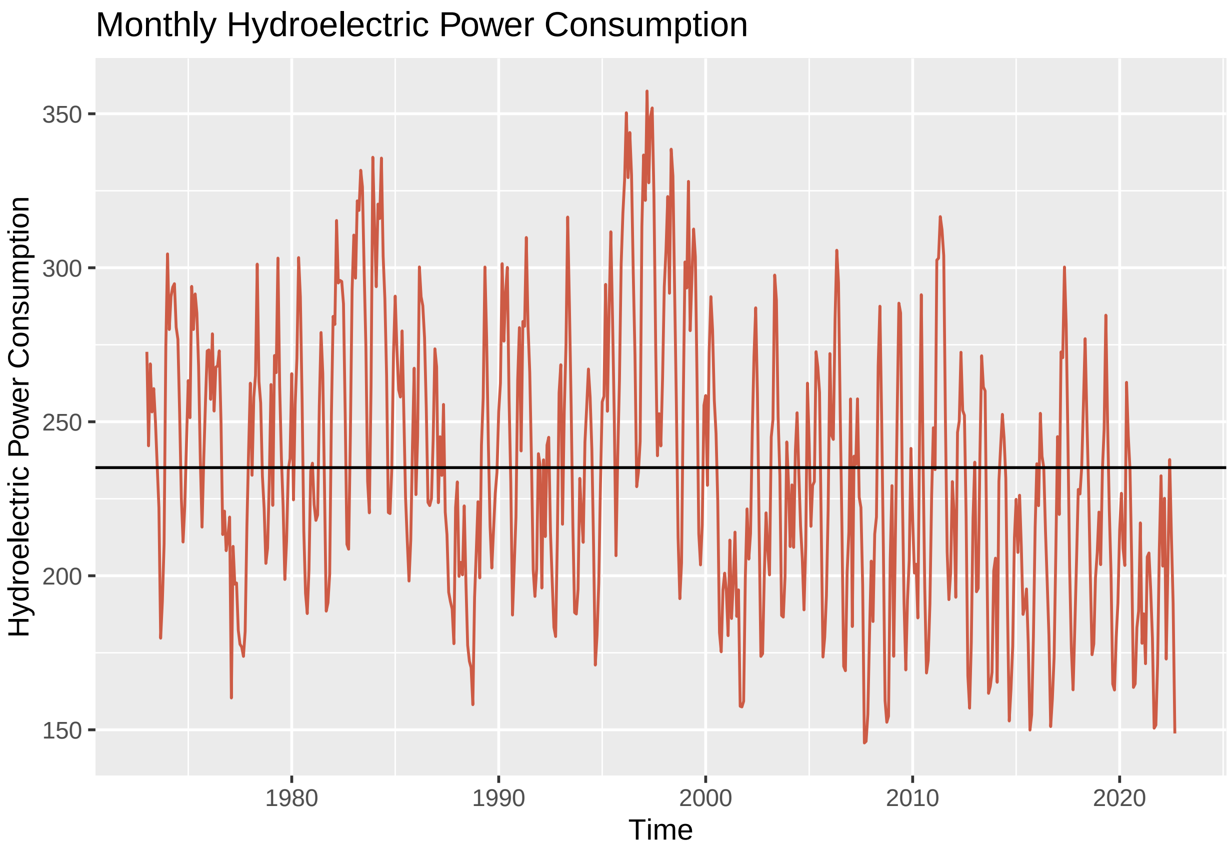
```r
        ylab = "Energy Biomass Production") +
        geom_hline(yintercept = Biomass_mean)
```




```r
#Total Renewable Energy Production
autoplot(ts_energy_data[,2],main = "Monthly Renewable Energy Production",
        col = "chartreuse4",
        ylab = "Renewable Energy Production") +
        geom_hline(yintercept = Renewable_mean)
```




```r
#Hydroelectric Power Consumption
autoplot(ts_energy_data[,3],main = "Monthly Hydroelectric Power Consumption",
         col = "coral3",
         ylab = "Hydroelectric Power Consumption") +
         geom_hline(yintercept = Hydroelect_mean)
```



Question 5

Compute the correlation between these three series. Are they significantly correlated? Explain your answer. >Answer: A correlation value near 1 or -1 indicates a strong positive or negative correlation. Since the correlation between total biomass and renewable energy production is 0.919, there is a strong positive correlation between the two. In contrast, hydroelectric power consumption and renewable energy production display a weak negative correlation at -0.0996. Hydroelectric power consumption and biomass energy production also show a relatively weak negative correlation at -0.210.

```
cor(ts_energy_data[, c(1, 2, 3)])
```

```
##                                    Total.Biomass.Energy.Production
## Total.Biomass.Energy.Production                          1.0000000
## Total.Renewable.Energy.Production                        0.9185941
## Hydroelectric.Power.Consumption                         -0.2998201
##                                    Total.Renewable.Energy.Production
## Total.Biomass.Energy.Production                           0.91859411
## Total.Renewable.Energy.Production                         1.00000000
## Hydroelectric.Power.Consumption                          -0.09958758
##                                    Hydroelectric.Power.Consumption
## Total.Biomass.Energy.Production                        -0.29982013
## Total.Renewable.Energy.Production                      -0.09958758
## Hydroelectric.Power.Consumption                         1.00000000
```

```
#Correlation between Biomass Energy Production and Renewable Energy Production
#p-value < 2.2e-16, reject the null hypothesis
cor.test(ts_energy_data[, 1], ts_energy_data[,2])
```

```
##
##  Pearson's product-moment correlation
##
## data:  ts_energy_data[, 1] and ts_energy_data[, 2]
## t = 56.697, df = 595, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.9050636 0.9302668
## sample estimates:
##       cor
## 0.9185941
```

```
#Correlation between Biomass Energy Production and Hydroelectric Power Consumption
#p-value = 7.256e-14, reject the null hypothesis
cor.test(ts_energy_data[, 1], ts_energy_data[,3])
```

```
##
##  Pearson's product-moment correlation
##
## data:  ts_energy_data[, 1] and ts_energy_data[, 3]
## t = -7.6661, df = 595, p-value = 7.256e-14
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.3711363 -0.2249878
## sample estimates:
##        cor
## -0.2998201
```

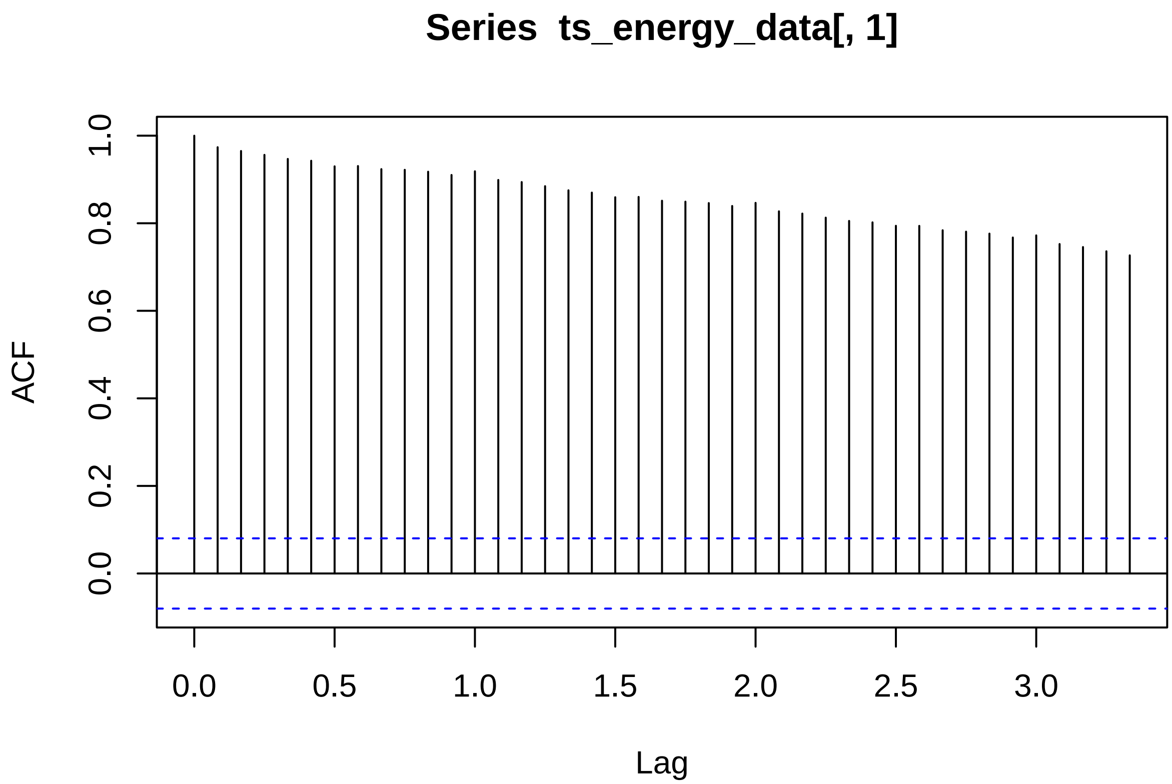
```
#Correlation between BRenewable Energy Production and Hydroelectric Power Consumption
#p p-value = 0.01492, reject the null hypothesis
cor.test(ts_energy_data[, 2], ts_energy_data[,3])
```

```
##
##  Pearson's product-moment correlation
##
## data:  ts_energy_data[, 2] and ts_energy_data[, 3]
## t = -2.4413, df = 595, p-value = 0.01492
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.17840723 -0.01949801
## sample estimates:
##         cor
## -0.09958758
```

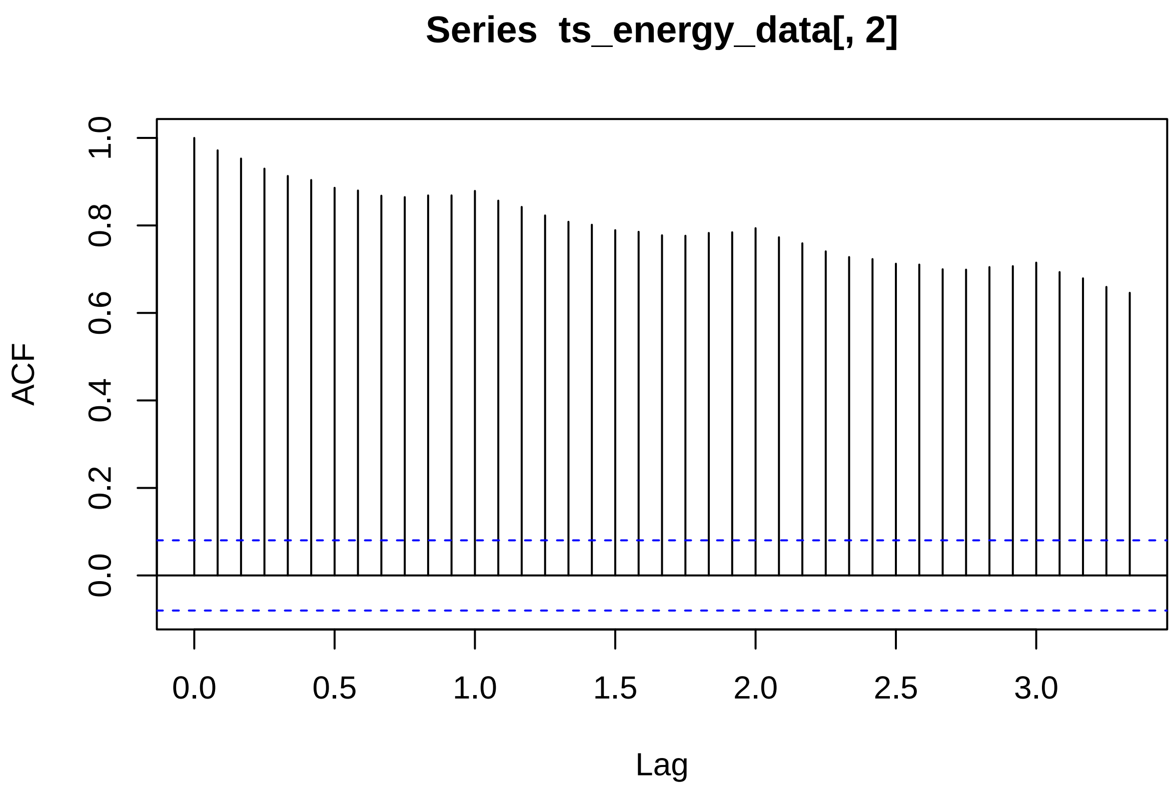
Question 6

Compute the autocorrelation function from lag 1 up to lag 40 for these three variables. What can you say about these plots? Do the three of them have the same behavior? > Answer: The biomass energy production and renewable energy production graph behave similarily with significant, positive acf values (i.e., high correlation) as seen in the plots displayed below. In contrast, the hydroelectric power consumption behaves differently with more of an oscillating pattern which can indicate seasonality.

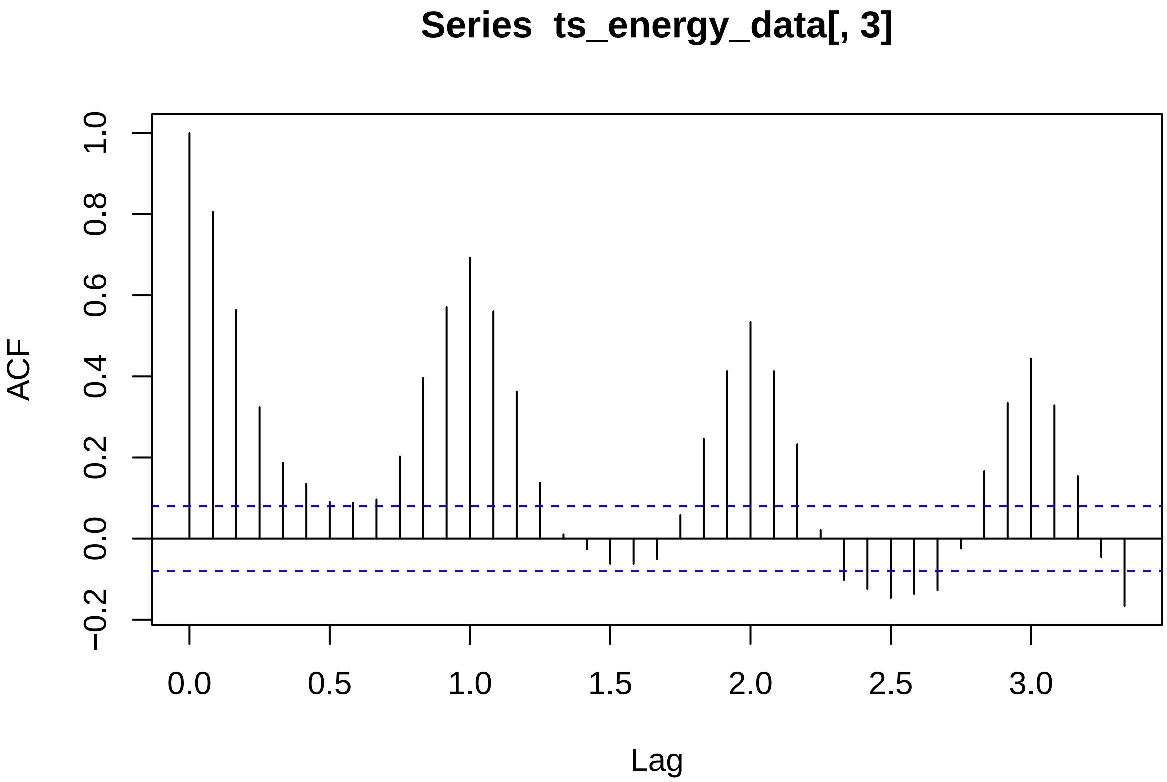
```
#Total Biomass Energy Production
acf(ts_energy_data[,1],lag.max=40, type="correlation",plot=TRUE)
```




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



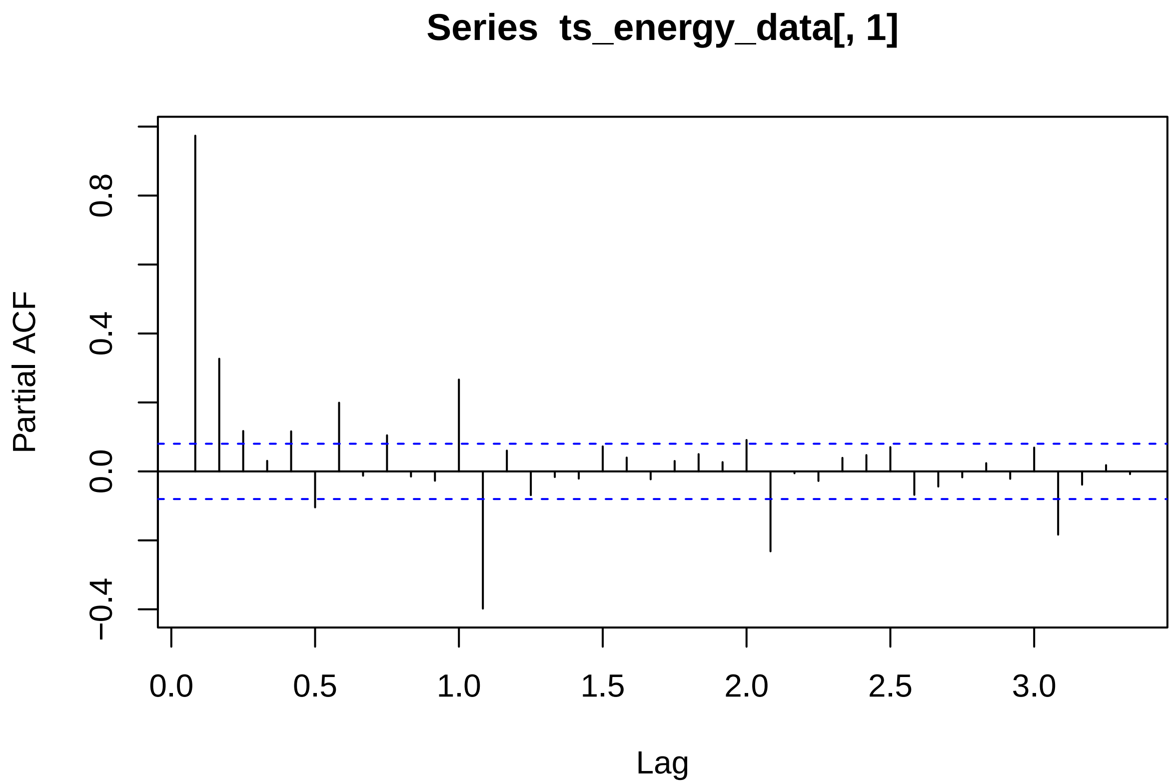

```r
#Hydroelectric Power Consumption
acf(ts_energy_data[,3],lag.max=40, type="correlation", plot=TRUE)
```



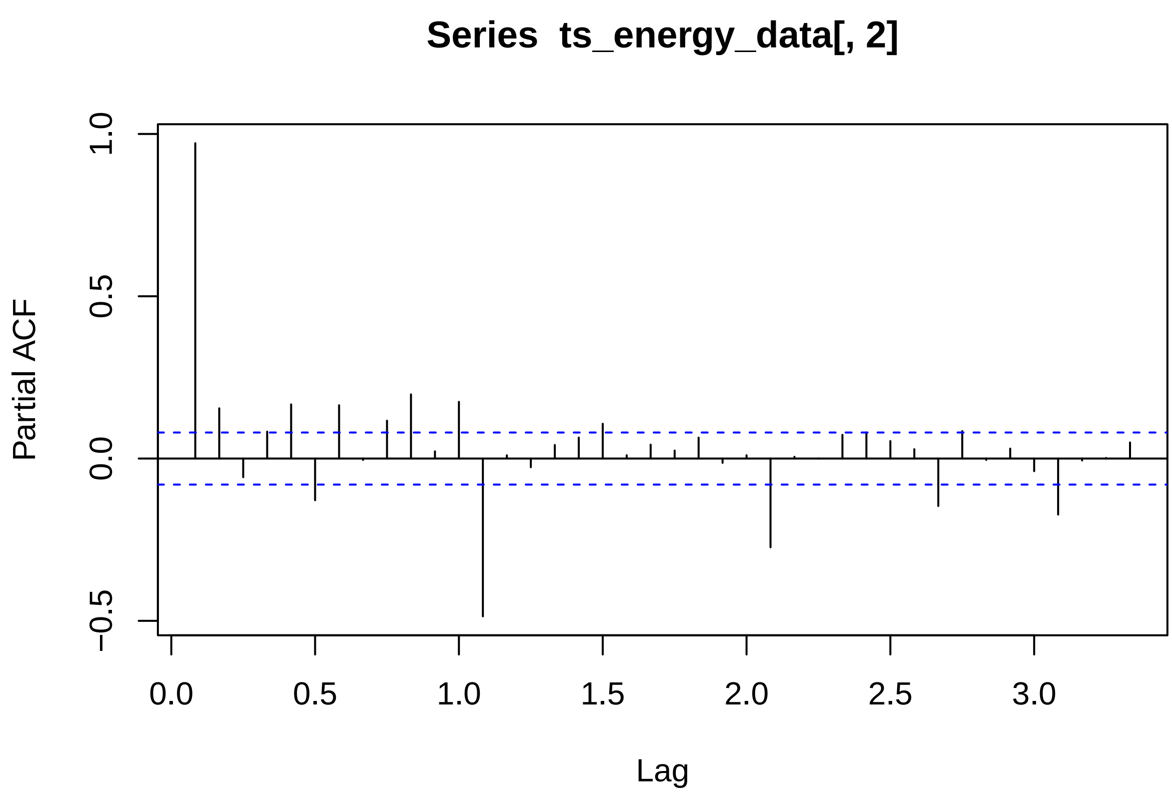
Question 7

Compute the partial autocorrelation function from lag 1 to lag 40 for these three variables. How these plots differ from the ones in Q6? > Answer: The values displayed on these plots are far more muted with less significant correlation and show a different pattern when compared to the acf plots since intermediate value influences are omitted.

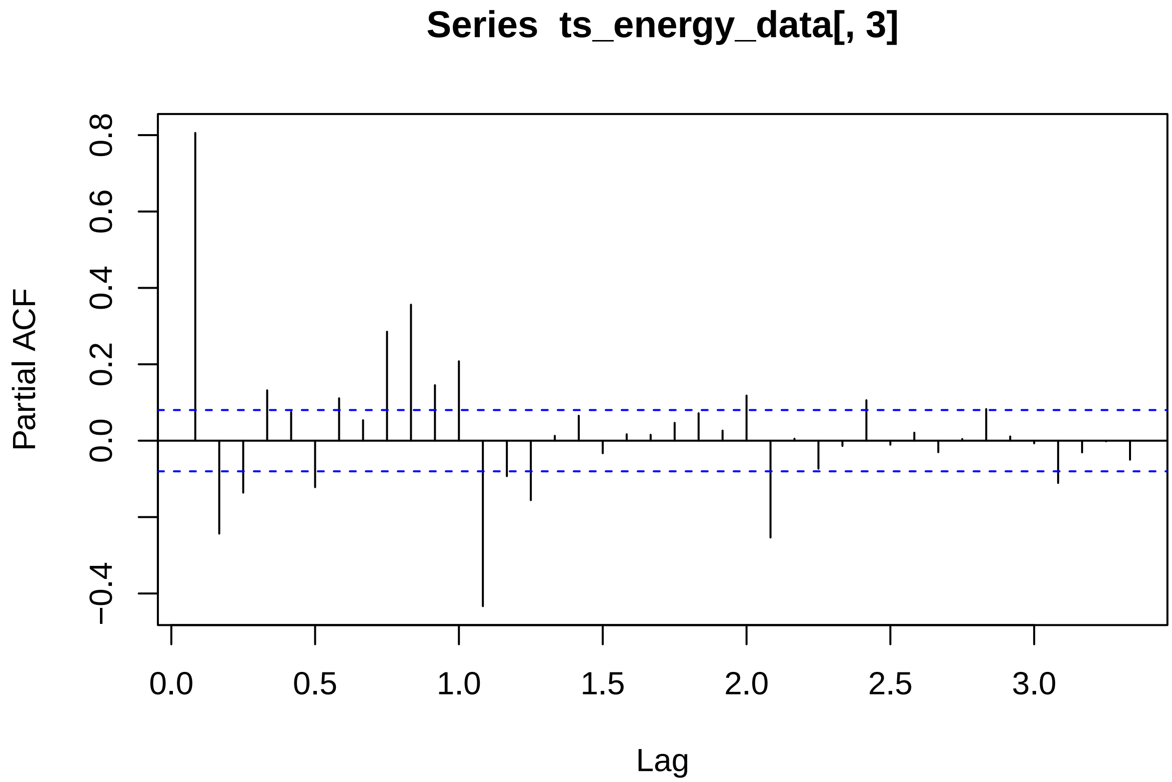
```
#Total Biomass Energy Production
pacf(ts_energy_data[,1],lag.max=40, plot=TRUE)
```




```
#Total Renewable Energy Production
pacf(ts_energy_data[,2],lag.max=40, plot=TRUE)
```




```
#Hydroelectric Power Consumption
pacf(ts_energy_data[,3],lag.max=40, plot=TRUE)
```


Series ts_energy_data[, 3]
Partial ACF
0.8
0.6
0.4
0.2
0.0
−0.4
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Lag