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

Assignment 2 - Due date 01/27/26

Maeve Gualtieri-Reed

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_Sp26.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.\

Data set information

Consider the data provided in the spreadsheet "Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source on our **Data** folder. The data comes from the US Energy Information and Administration and corresponds to the December 2025 Monthly Energy Review. The spreadsheet is ready to be used. Refer to the file "M2_ImportingData_XLSX.Rmd" in our Lessons folder for instructions on how to read *.xlsx* files.

```
#Importing data set
energy_data1 <- read_excel("/home/guest/TSA_Sp26/Data/Table_10.1_Renewable_Energy_Production_and_Consump

#Now let's extract the column names from row 11
read_col_names <- read_excel("/home/guest/TSA_Sp26/Data/Table_10.1_Renewable_Energy_Production_and_Consu

#Assign the column names to the data set
colnames(energy_data1) <- read_col_names

#make date column a date data type
energy_data1$Month <- as.Date(ymd(energy_data1$Month))
```

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.

```
bio_renew_hydro <- energy_data1 %>%
  select('Month',
         'Total Biomass Energy Production',
         'Total Renewable Energy Production',
         'Hydroelectric Power Consumption')

head(bio_renew_hydro)
```

```
## # A tibble: 6 x 4
##   Month      `Total Biomass Energy Production` `Total Renewable Energy Producti~1
##   <date>                                 <dbl>                              <dbl>
## 1 1973-01-01                              130.                               220.
## 2 1973-02-01                              117.                               197.
## 3 1973-03-01                              130.                               219.
## 4 1973-04-01                              126.                               209.
## 5 1973-05-01                              130.                               216.
## 6 1973-06-01                              126.                               208.
## # i abbreviated name: 1: `Total Renewable Energy Production`
## # i 1 more variable: `Hydroelectric Power Consumption` <dbl>
```

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_energy <- ts(bio_renew_hydro[,2:4],
                    start=c(1973,1),
                    frequency=12)
ts_energy
```

Question 3

Compute mean and standard deviation for these three series.

```
mean_biomass <- mean(ts_energy[,1])
mean_renew <- mean(ts_energy[,2])
mean_hydro <- mean(ts_energy[,3])

mean_biomass
```

```
## [1] 286.0489
```

```
mean_renew
```

```
## [1] 409.1952
```

```
mean_hydro
```

```
## [1] 79.35682
```

```
sd_biomass <- sd(ts_energy[,1])
sd_renew <- sd(ts_energy[,2])
sd_hydro <- sd(ts_energy[,3])

sd_biomass
```

```
## [1] 96.21209
```

```
sd_renew
```

```
## [1] 151.4223
```

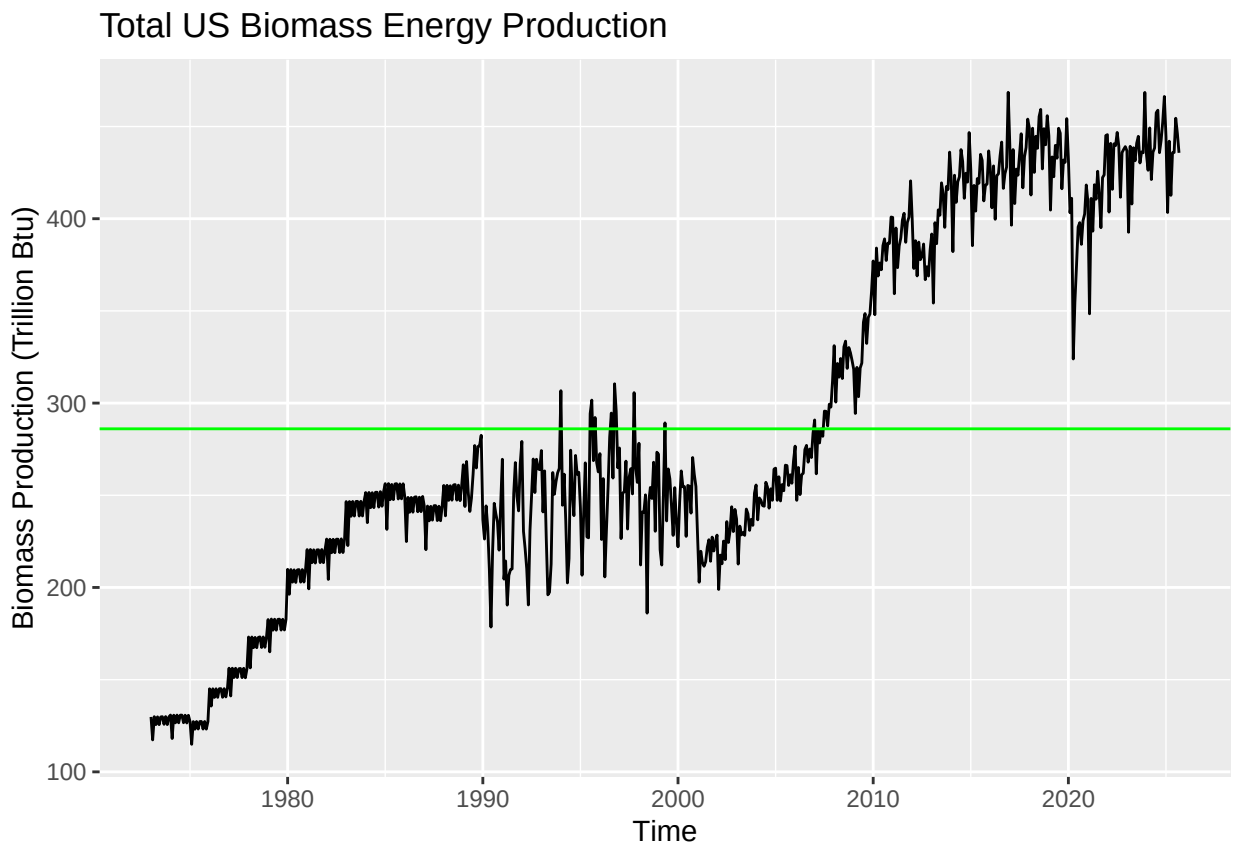
```
sd_hydro
```

```
## [1] 14.1202
```

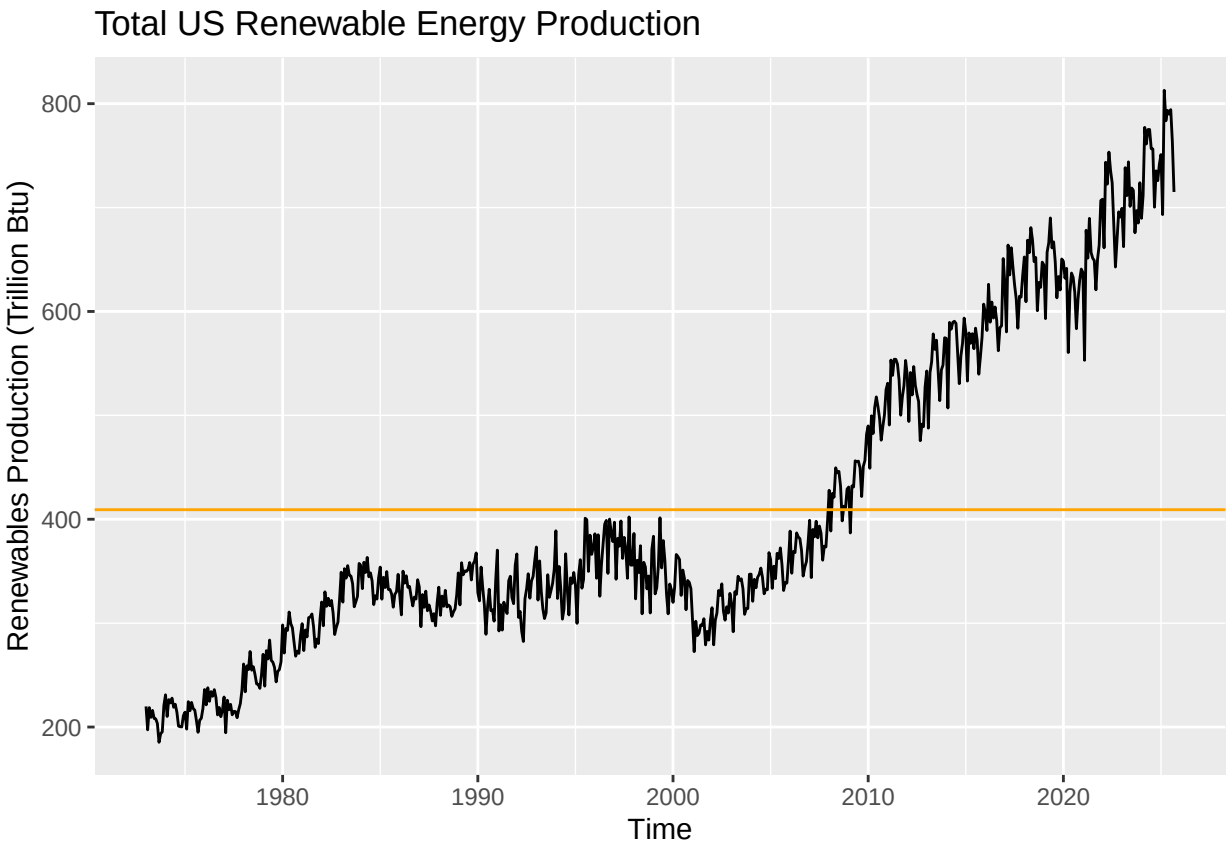
Question 4

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.

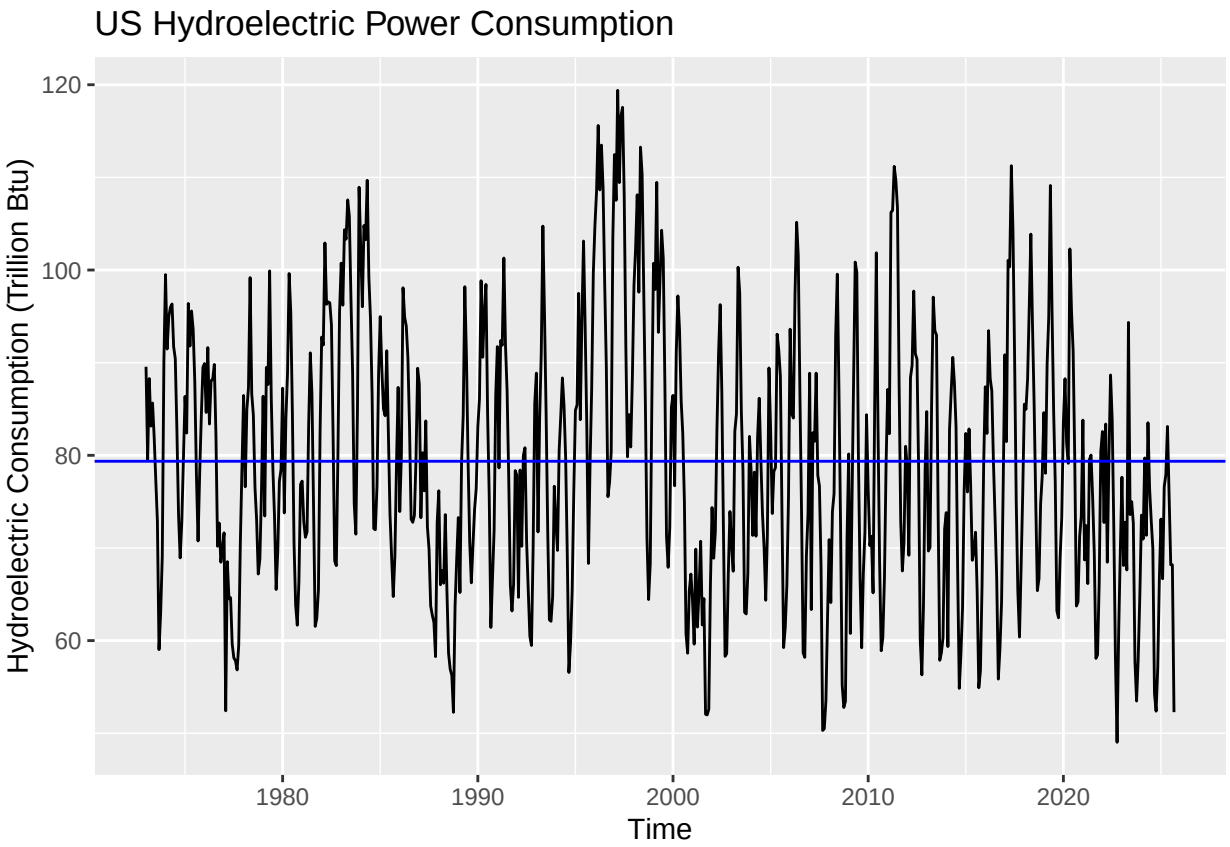
```
#Biomass TS plot
autoplot(ts_energy[,1]) +
  xlab("Time") +
  ylab("Biomass Production (Trillion Btu)")+
  ggtitle("Total US Biomass Energy Production")+
  geom_hline(yintercept = mean_biomass, color = "green")
```




```
#Renewables TS plot
autoplot(ts_energy[,2]) +
  xlab("Time") +
  ylab("Renewables Production (Trillion Btu)")+
  ggtitle("Total US Renewable Energy Production")+
  geom_hline(yintercept = mean_renew, color = "orange")
```




```
#Hydro TS plot
autoplot(ts_energy[,3]) +
  xlab("Time") +
  ylab("Hydroelectric Consumption (Trillion Btu)") +
  ggtitle("US Hydroelectric Power Consumption") +
  geom_hline(yintercept = mean_hydro, color = "blue")
```



Question 5

Compute the correlation between these three series. Are they significantly correlated? Explain your answer.

```
cor(ts_energy)
```

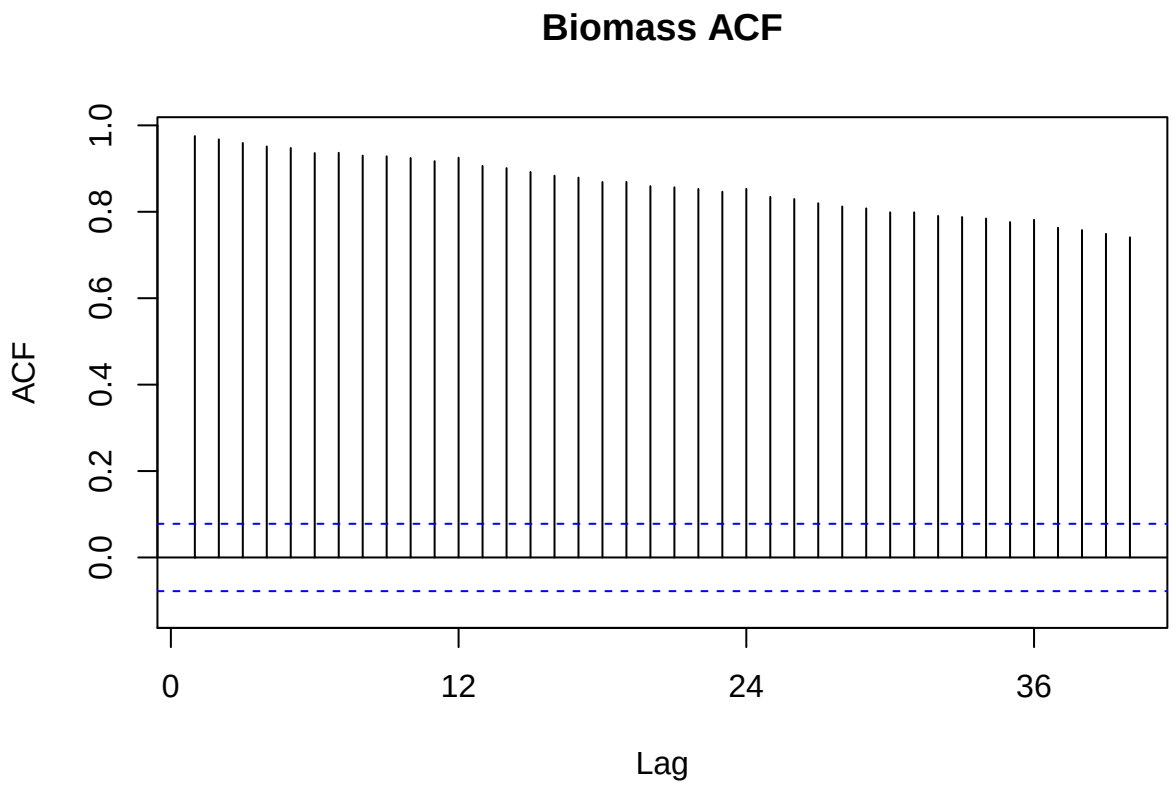
```
##                                  Total Biomass Energy Production
## Total Biomass Energy Production                        1.0000000
## Total Renewable Energy Production                      0.9652985
## Hydroelectric Power Consumption                       -0.1347374
##                                  Total Renewable Energy Production
## Total Biomass Energy Production                         0.96529851
## Total Renewable Energy Production                       1.00000000
## Hydroelectric Power Consumption                        -0.05842436
##                                  Hydroelectric Power Consumption
## Total Biomass Energy Production                       -0.13473742
## Total Renewable Energy Production                     -0.05842436
## Hydroelectric Power Consumption                        1.00000000
```

The correlation matrix shows a strong positive correlation between renewables and biomass (0.965). There does not seem to be a significant correlation between hydro and renewables or biomass (-0.058 and -0.135 respectively). This is supported by the time series plots in question 4 as renewables and biomass seem to be increasing at a similar rate while hydro has fluctuated but does not show and overall trend.

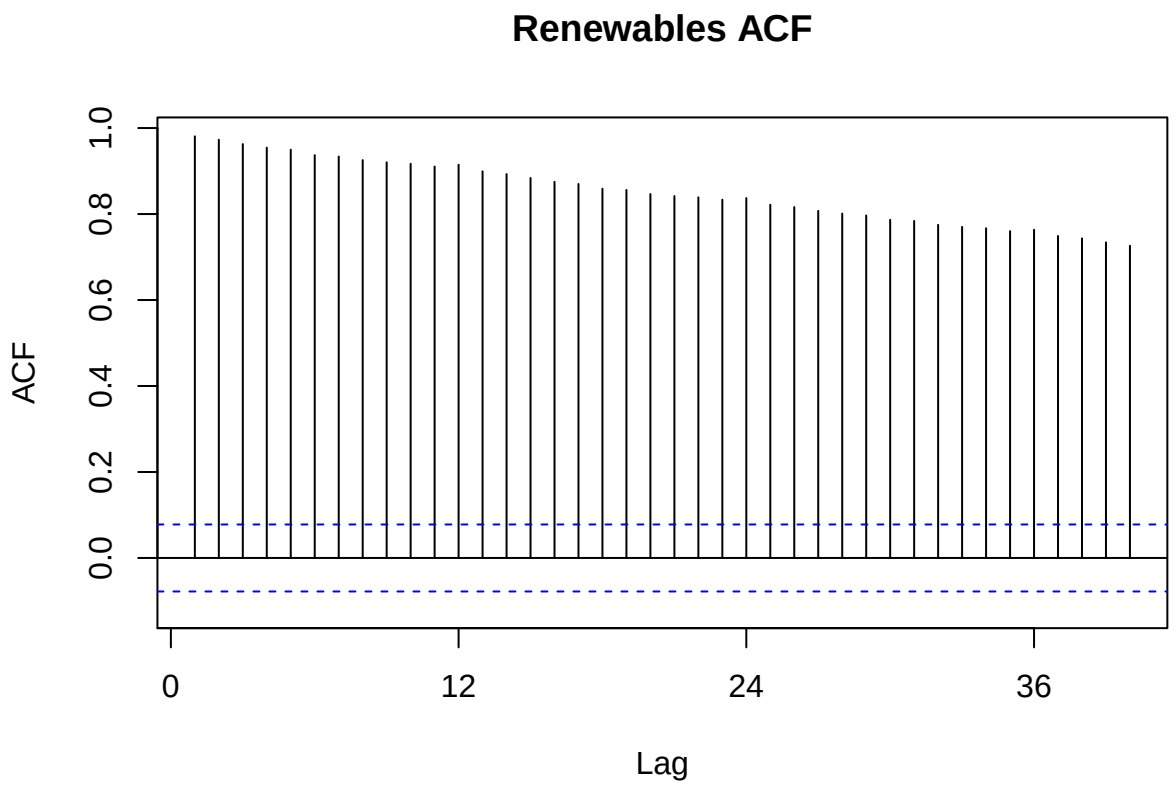
Question 6

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?

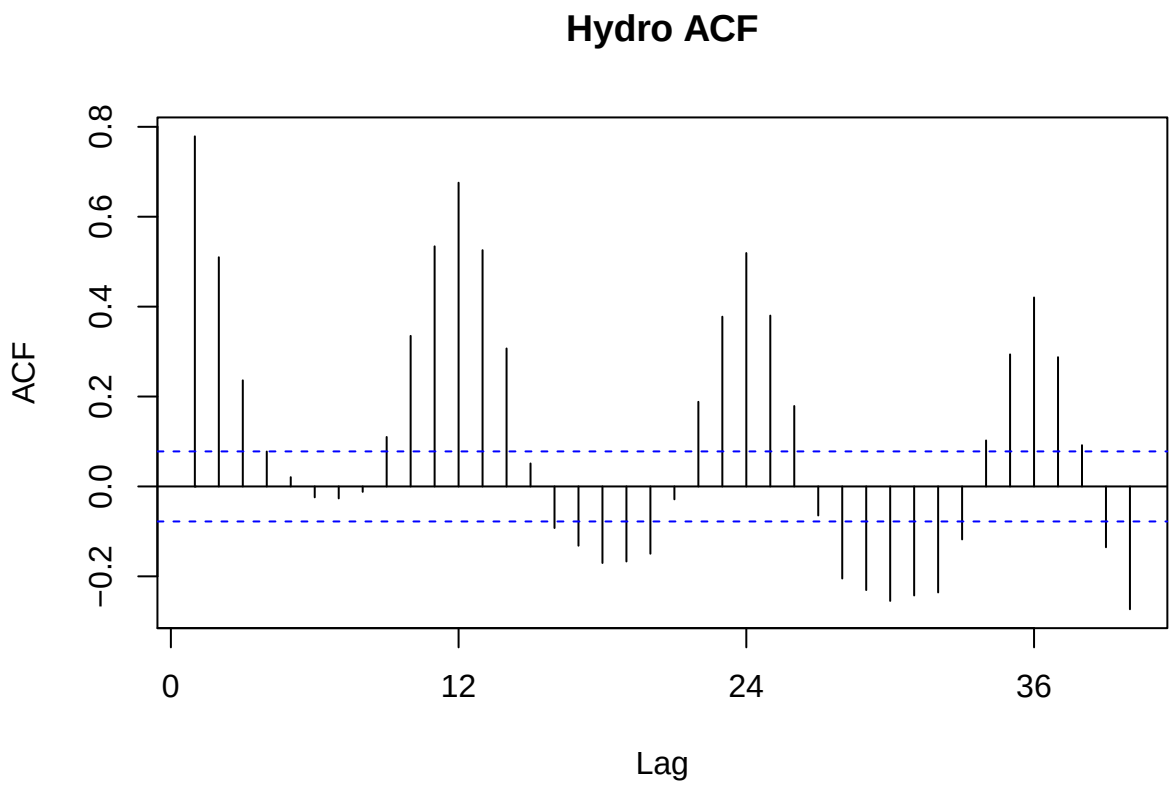
```
Biomass_acf= Acf(ts_energy[,1], lag.max = 40, main = "Biomass ACF")
```




```
renew_acf= Acf(ts_energy[,2], lag.max = 40, main = "Renewables ACF")
```




```
hydro_acf= Acf(ts_energy[,3], lag.max = 40, main = "Hydro ACF")
```

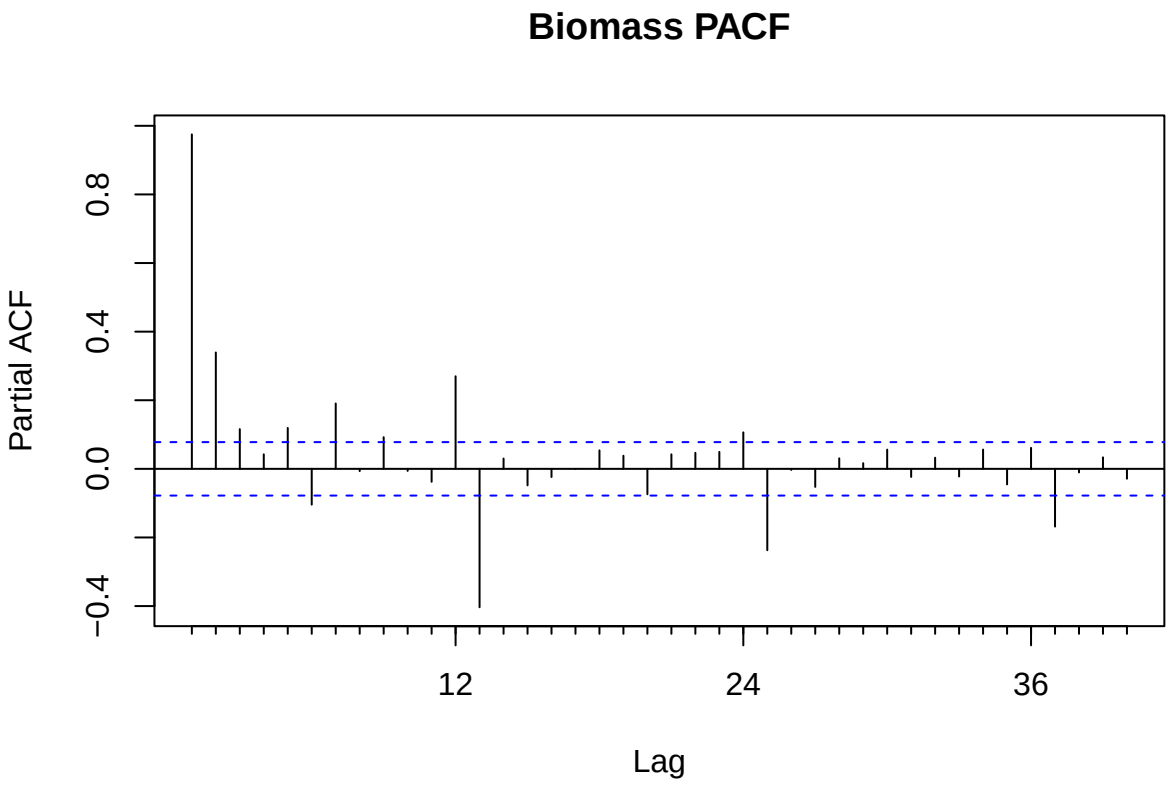


Biomass and renewables have a similar looking functions with a lot of positive autocorrelation even for the later lags. This indicates that these are long memory series. There is not a clear seasonal component, but this could be due to the clear increasing trend shown in the plots created for question 4. The hydro acf looks completely different with very clear seasonal fluctuations.

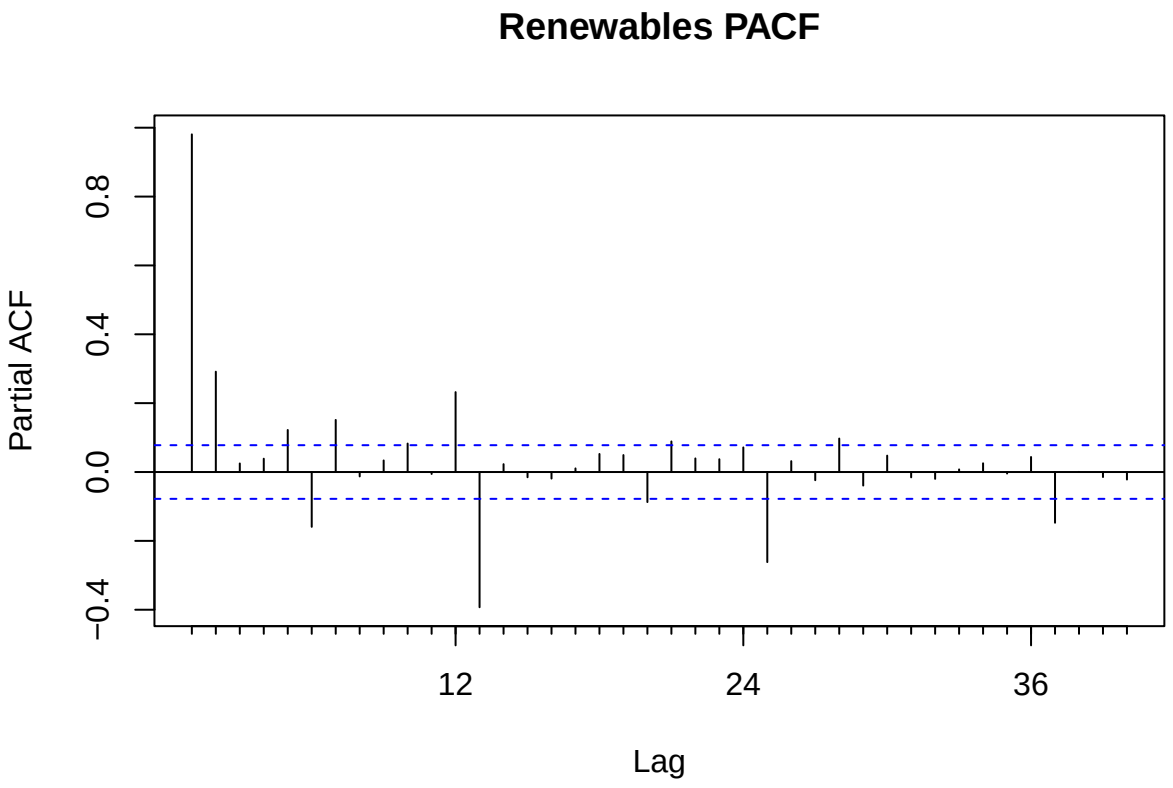
Question 7

Compute the partial autocorrelation function from lag 1 to lag 40 for these three variables. How these plots differ from the ones in Q6?

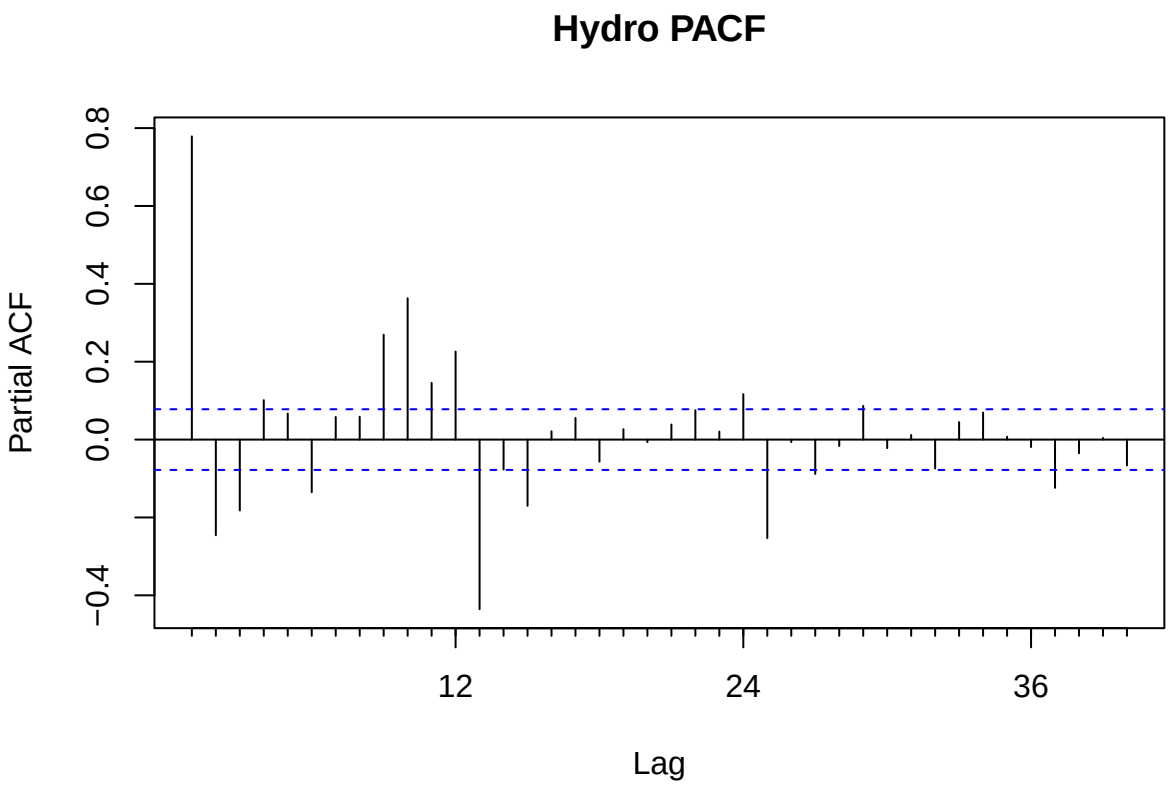
```
Biomass_pacf= Pacf(ts_energy[,1], lag.max = 40, main = "Biomass PACF")
```




```
renew_pacf= Pacf(ts_energy[,2], lag.max = 40, main = "Renewables PACF")
```




```
hydro_pacf= Pacf(ts_energy[,3], lag.max = 40, main = "Hydro PACF")
```



Especially the renewable and hydro PACF plots are much less extreme than their ACF plots. In renewable and hydro, there seems to be a significant variable around T-13 that should be further investigated. The hydro PACF plot shows that the first 12 lags seem to be having a significant affect, perhaps indicating that the annual water flow has an impact on monthly production and this is a long memory series.