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

Assignment 2 - Due date 01/27/26

Nicole Gutkowski

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.\

```r
# Load/install required package here

library(lubridate)
```

```
##
## Attaching package: 'lubridate'

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

```r
library(ggplot2)
library(forecast)
```

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

```r
library(readxl)
library(openxlsx)
library(cowplot)
```

```
##
## Attaching package: 'cowplot'

## The following object is masked from 'package:lubridate':
##
##     stamp
```

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.

```r
# Importing data set

energy_data1 = read_xlsx(path = "./Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source
    skip = 12, sheet = "Monthly Data", col_names = FALSE)
```

```
## New names:
## * `` -> `...1`
## * `` -> `...2`
## * `` -> `...3`
## * `` -> `...4`
## * `` -> `...5`
## * `` -> `...6`
## * `` -> `...7`
## * `` -> `...8`
## * `` -> `...9`
## * `` -> `...10`
## * `` -> `...11`
## * `` -> `...12`
## * `` -> `...13`
## * `` -> `...14`
```

```r
# Now let's extract the column names from row 11
read_col_names <- read_excel(path = "./Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_So
    skip = 10, n_max = 1, sheet = "Monthly Data", col_names = FALSE)
```

```
## New names:
## * `` -> `...1`
## * `` -> `...2`
## * `` -> `...3`
## * `` -> `...4`
## * `` -> `...5`
## * `` -> `...6`
## * `` -> `...7`
```

```
## * `` -> `...8`
## * `` -> `...9`
## * `` -> `...10`
## * `` -> `...11`
## * `` -> `...12`
## * `` -> `...13`
## * `` -> `...14`
```

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

# Visualize the first rows of the data set
head(energy_data1)
```

```
## # A tibble: 6 x 14
##   Month               `Wood Energy Production` `Biofuels Production`
##   <dttm>                                 <dbl> <chr>
## 1 1973-01-01 00:00:00                     130. Not Available
## 2 1973-02-01 00:00:00                     117. Not Available
## 3 1973-03-01 00:00:00                     130. Not Available
## 4 1973-04-01 00:00:00                     125. Not Available
## 5 1973-05-01 00:00:00                     130. Not Available
## 6 1973-06-01 00:00:00                     125. Not Available
## # i 11 more variables: `Total Biomass Energy Production` <dbl>,
## #   `Total Renewable Energy Production` <dbl>,
## #   `Hydroelectric Power Consumption` <dbl>,
## #   `Geothermal Energy Consumption` <dbl>, `Solar Energy Consumption` <chr>,
## #   `Wind Energy Consumption` <chr>, `Wood Energy Consumption` <dbl>,
## #   `Waste Energy Consumption` <dbl>, `Biofuels Consumption` <chr>,
## #   `Total Biomass Energy Consumption` <dbl>, ...
```

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.

```r
subset_df = energy_data1[, c("Total Biomass Energy Production",
    "Total Renewable Energy Production", "Hydroelectric Power Consumption")]

head(subset_df)
```

```
## # A tibble: 6 x 3
##   Total Biomass Energy Productio~1 Total Renewable Ener~2 Hydroelectric Power ~3
##                              <dbl>                  <dbl>                  <dbl>
## 1                             130.                   220.                   89.6
## 2                             117.                   197.                   79.5
## 3                             130.                   219.                   88.3
## 4                             126.                   209.                   83.2
## 5                             130.                   216.                   85.6
## 6                             126.                   208.                   82.1
## # i abbreviated names: 1: `Total Biomass Energy Production`,
```

```
## #   2: ‘Total Renewable Energy Production‘,
## #   3: ‘Hydroelectric Power Consumption‘
```

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_data = ts(data = subset_df, start = c(1, 1))
```

Question 3

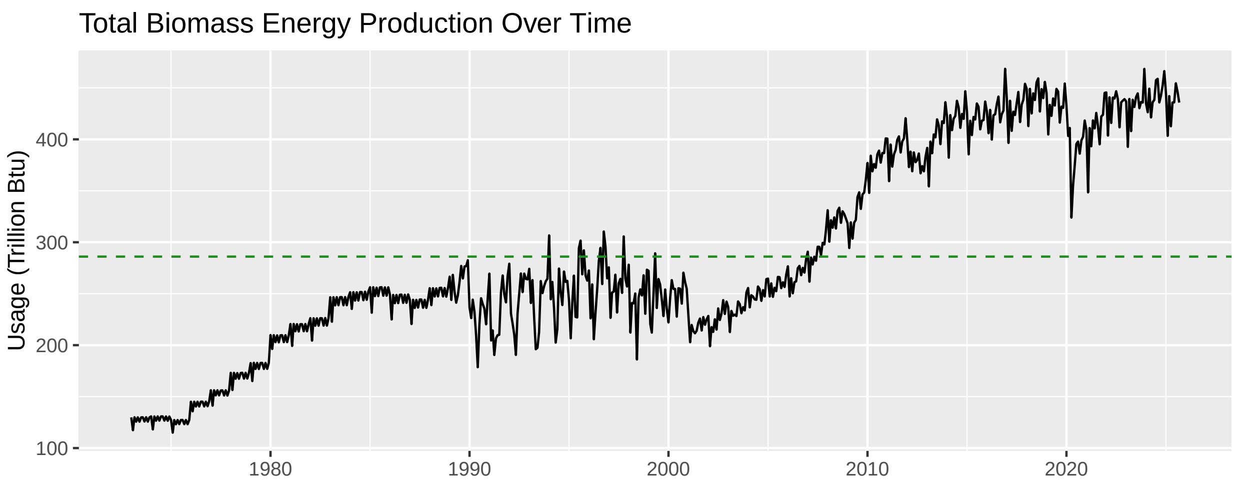
Compute mean and standard deviation for these three series.

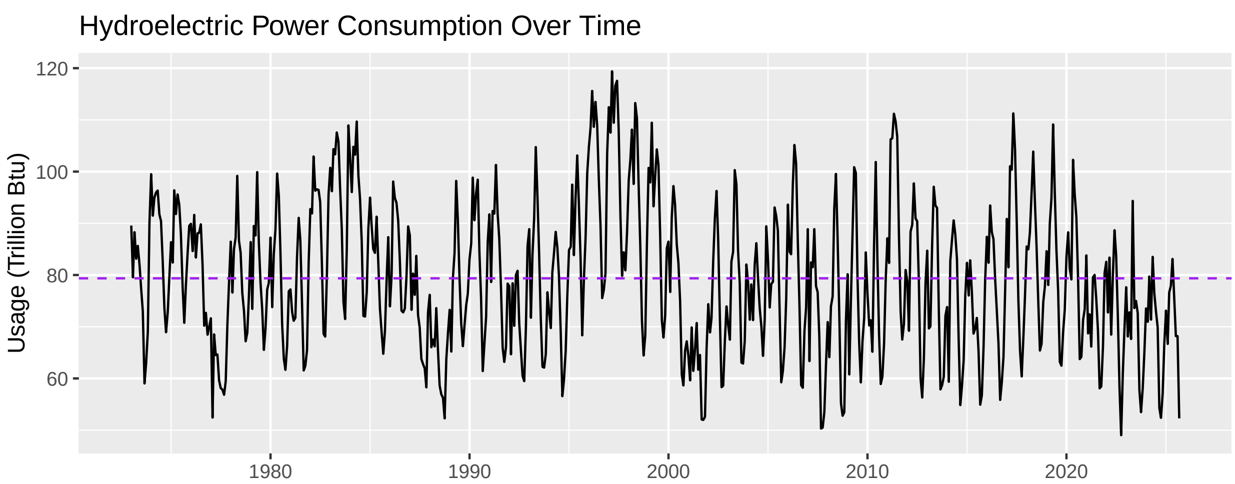
```
mean1 = mean(ts_data[, 1])
mean2 = mean(ts_data[, 2])
mean3 = mean(ts_data[, 3])

std1 = sd(ts_data[, 1])
std2 = sd(ts_data[, 2])
std3 = sd(ts_data[, 3])
```

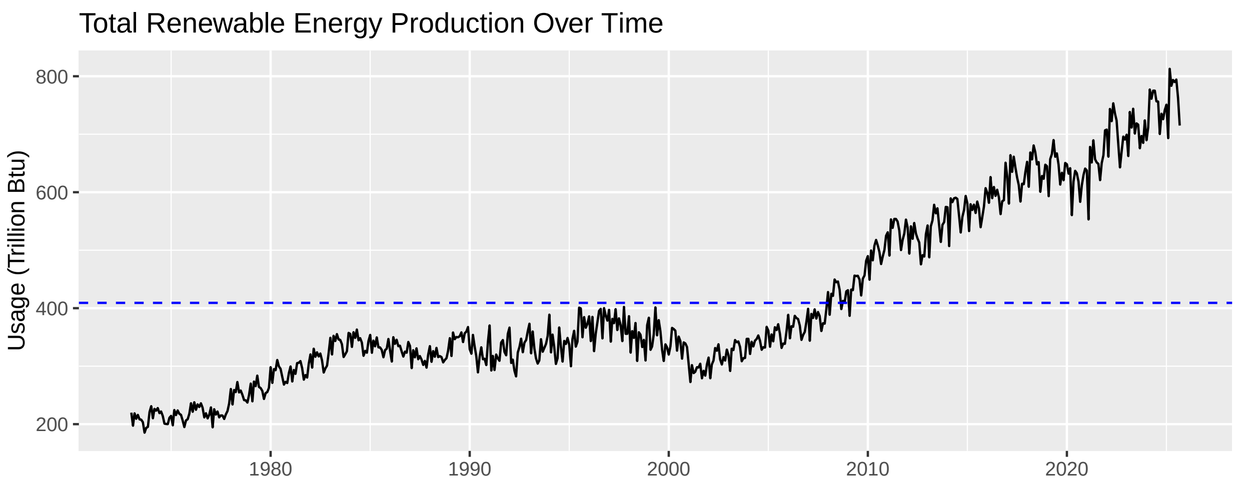
Question 4

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_data = energy_data1[, c("Month", "Total Biomass Energy Production")]
renew_data = energy_data1[, c("Month", "Total Renewable Energy Production")]
hydro_data = energy_data1[, c("Month", "Hydroelectric Power Consumption")]

biomass_plot <- ggplot(biomass_data, aes(x = Month, y = `Total Biomass Energy Production`)) +
    geom_line() + xlab("") + ylab("Usage (Trillion Btu)") + labs(title = "Total Biomass Energy Producti
    geom_hline(yintercept = mean(energy_data1$`Total Biomass Energy Production`),
        color = "forestgreen", linetype = "dashed")


renew_plot <- ggplot(renew_data, aes(x = Month, y = `Total Renewable Energy Production`)) +
    geom_line() + xlab("") + ylab("Usage (Trillion Btu)") + labs(title = "Total Renewable Energy Produc
    geom_hline(yintercept = mean(energy_data1$`Total Renewable Energy Production`),
        color = "blue", linetype = "dashed")

hydro_plot <- ggplot(hydro_data, aes(x = Month, y = `Hydroelectric Power Consumption`)) +
    geom_line() + xlab("") + ylab("Usage (Trillion Btu)") + labs(title = "Hydroelectric Power Consumpti
    geom_hline(yintercept = mean(energy_data1$`Hydroelectric Power Consumption`),
        color = "purple", linetype = "dashed")

plot_grid(biomass_plot, hydro_plot, renew_plot, ncol = 1)
```







Question 5

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

```r
correlation = cor(subset_df)

correlation
```

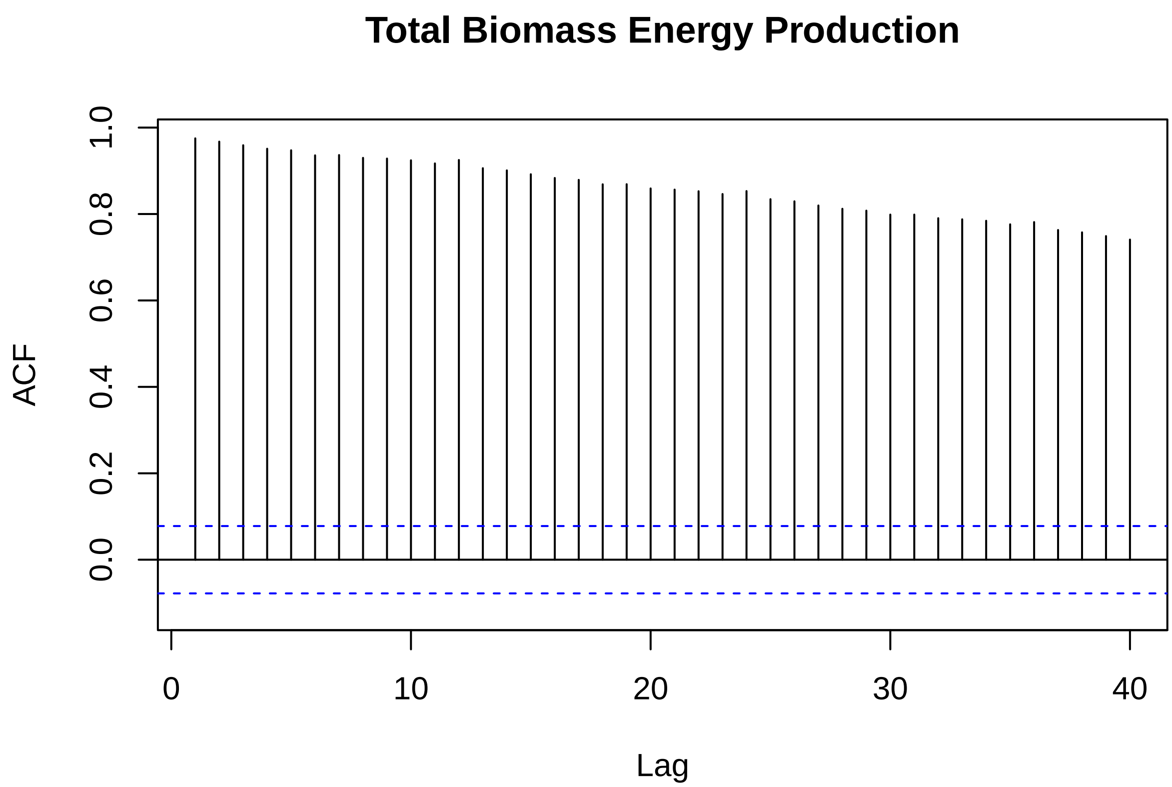
```
##                                 Total Biomass Energy Production
## Total Biomass Energy Production                       1.0000000
## Total Renewable Energy Production                     0.9652985
## Hydroelectric Power Consumption                      -0.1347374
##                                 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
```

Total Biomass energy production and total renewable energy production are highly correlated with each other (0.965), while hydroelectric power consumption is has little correlation with the other variables. Hydroelectric power consumption's correlations were -0.135 with total biomass energy production and -0.058 with total renewable energy production.

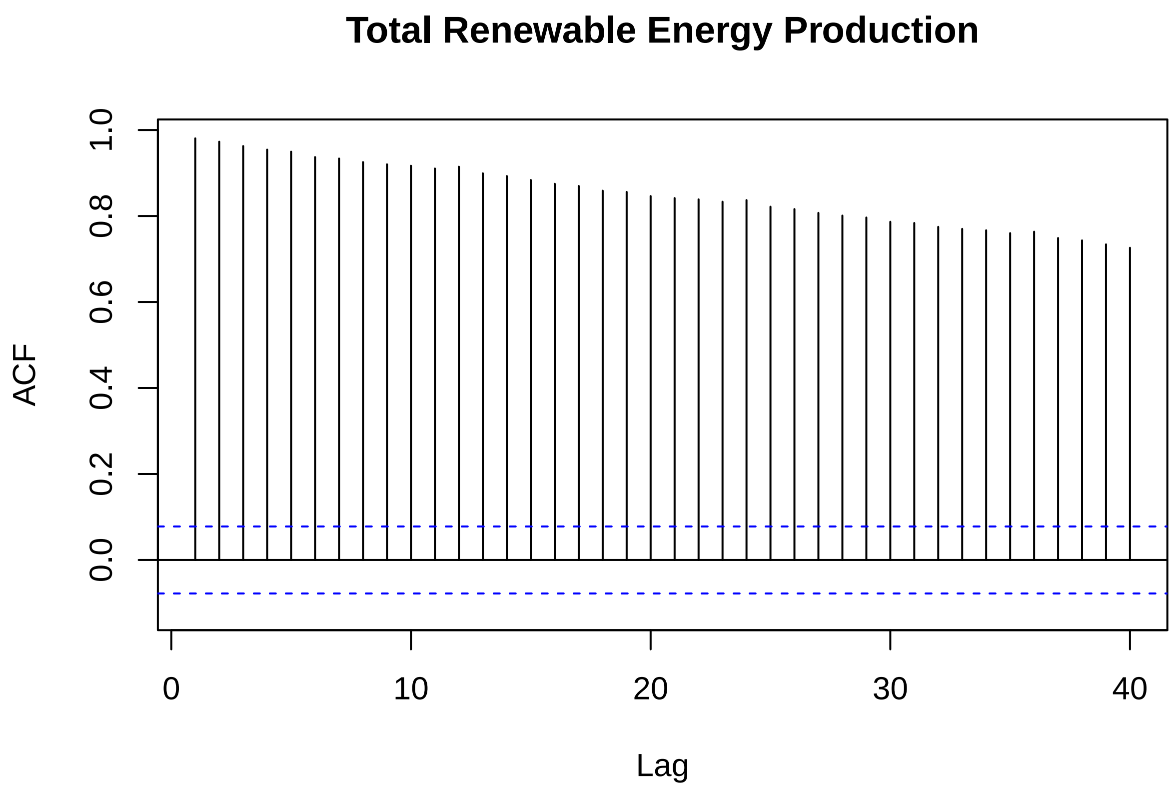
Question 6

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?

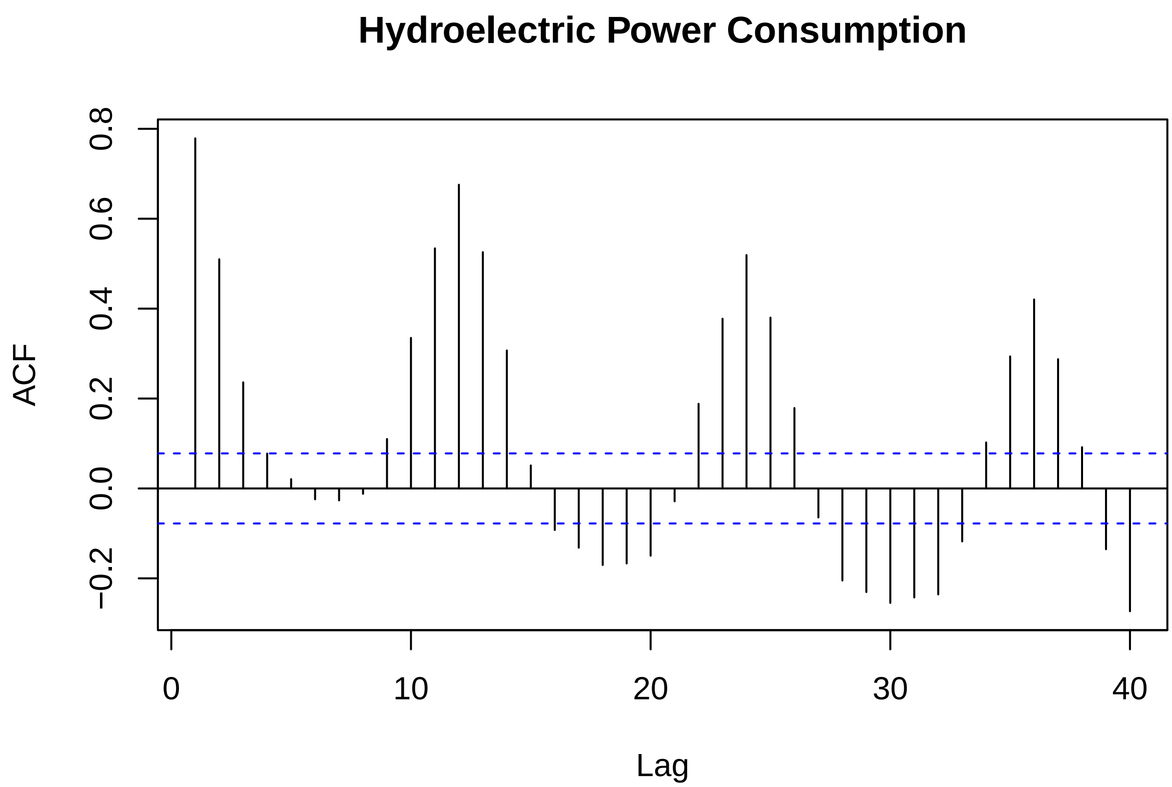
```r
ts1_plot = Acf(ts_data[, 1], lag.max = 40, main = "Total Biomass Energy Production")
```




```
ts1_acf = ts1_plot$acf

ts2_plot = Acf(ts_data[, 2], lag.max = 40, main = "Total Renewable Energy Production")
```




```
ts2_acf = ts2_plot$acf

ts3_plot = Acf(ts_data[, 3], lag.max = 40, main = "Hydroelectric Power Consumption")
```



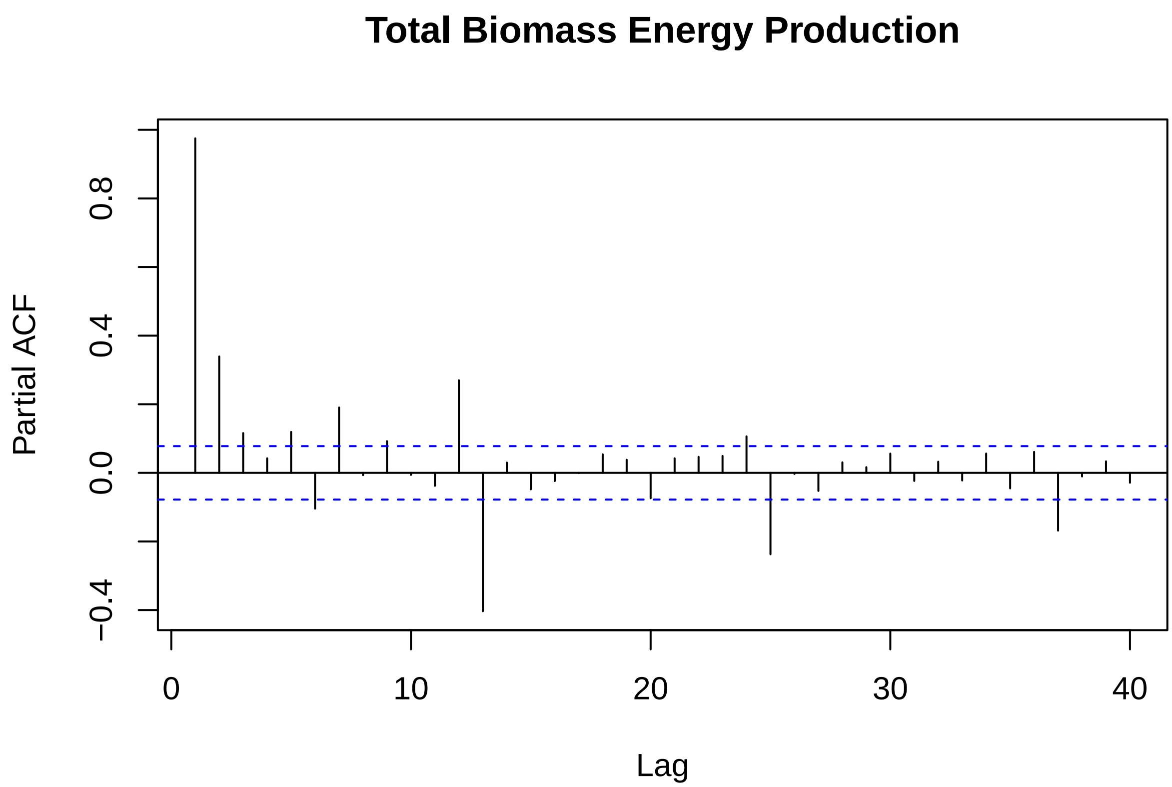

```r
ts3_acf = ts3_plot$acf
```

While Total Biomass Production and Total Renewable Energy Production follow the same general decreasing behavior, Hydroelectric Power Consumption follows a strong seasonal trend.

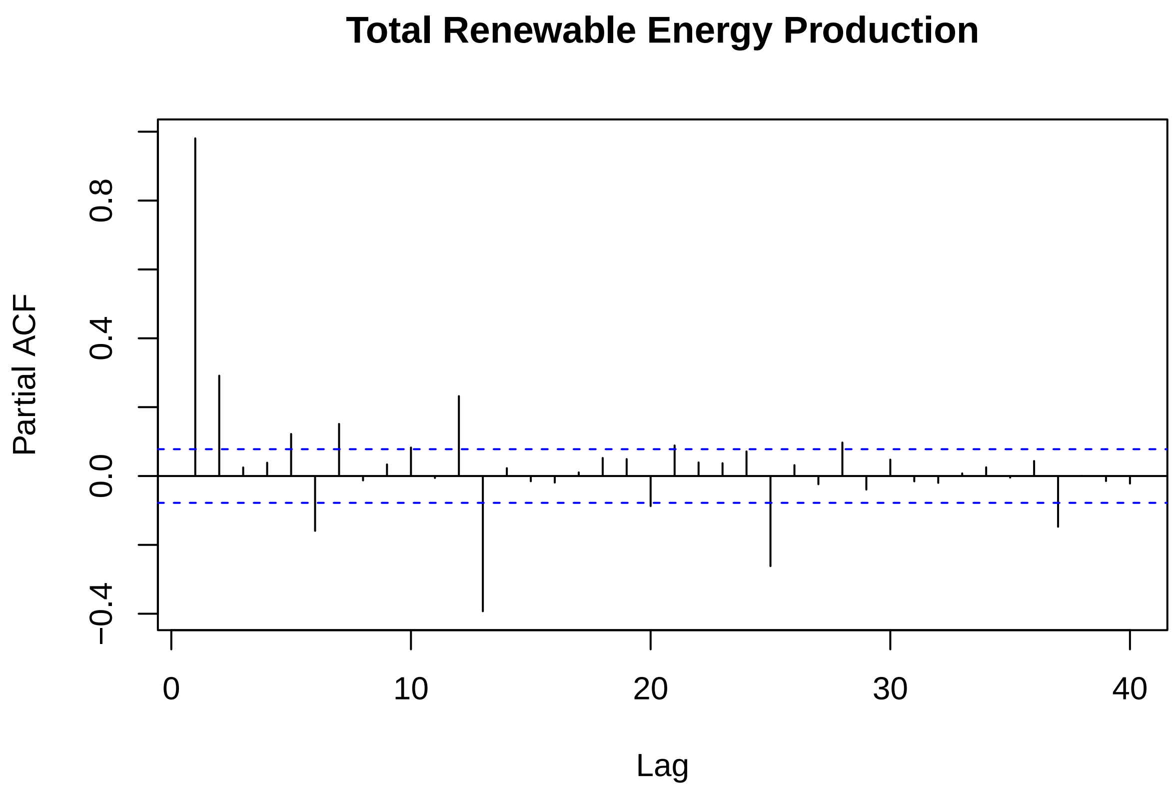
Question 7

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

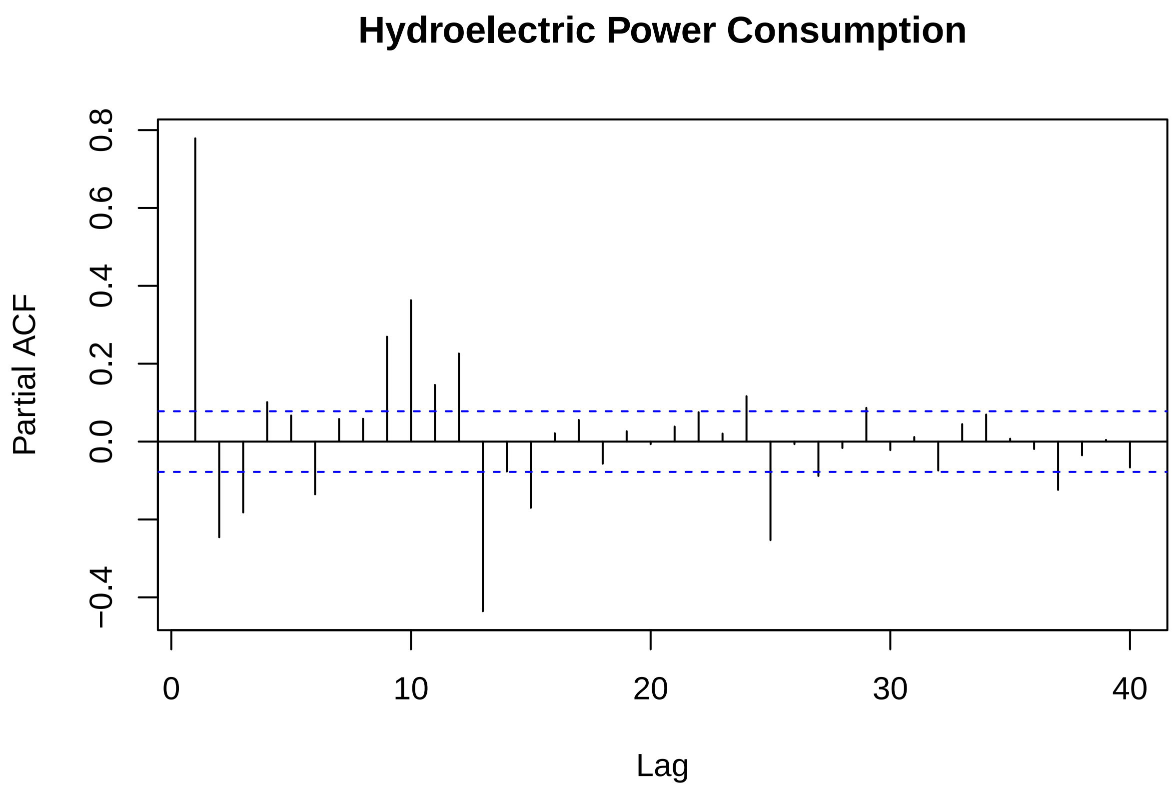
```r
ts1_pacf_plot = Pacf(ts_data[, 1], lag.max = 40, main = "Total Biomass Energy Production")
```




```
ts1_pacf = ts1_pacf_plot$acf

ts2_pacf_plot = Pacf(ts_data[, 2], lag.max = 40, main = "Total Renewable Energy Production")
```




```
ts2_pacf = ts2_pacf_plot$acf

ts3_pacf_plot = Pacf(ts_data[, 3], lag.max = 40, main = "Hydroelectric Power Consumption")
```




```
ts3_pacf = ts3_pacf_plot$acf
```

All of the partial autocorrelation plots show potential seasonal patterns. However, while partial autocorrelation decayed quickly for the biomass and total renewable energy plots, the hydroelectric partial autocorrelation plot showed stronger memory across the lags.