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

Assignment 2 - Due date 01/27/26

Matthew Fischer

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.\

```
#Load/install required package here
library(forecast)
```

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

```
library(tseries)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

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

```
library(ggplot2)
library("readxl")
```

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 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_data <- read_excel(path="../Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_Sourc
```

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

```r
read_col_names <- read_excel(path="../Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_Sou
```

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

```r
colnames(energy_data) <- read_col_names

head(energy_data)
```

```
## # 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
energy_subset <- data.frame(
  Total_Biomass_Energy_Production =
    energy_data$"Total Biomass Energy Production",
  Total_Renewable_Energy_Production =
    energy_data$"Total Renewable Energy Production",
  Hydroelectric_Power_Consumption =
    energy_data$"Hydroelectric Power Consumption"
)

head(energy_subset)
```

```
##   Total_Biomass_Energy_Production Total_Renewable_Energy_Production
## 1                         129.787                           219.839
## 2                         117.338                           197.330
## 3                         129.938                           218.686
## 4                         125.636                           209.330
## 5                         129.834                           215.982
## 6                         125.611                           208.249
##   Hydroelectric_Power_Consumption
## 1                          89.562
## 2                          79.544
## 3                          88.284
## 4                          83.152
## 5                          85.643
## 6                          82.060
```

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().

```r
energy_ts <- ts(
  energy_subset,
  start = c(1973, 1),
  frequency = 12
)
```

Question 3

Compute mean and standard deviation for these three series.

```r
mean_Total_Biomass_Energy_Production <- mean(energy_ts[, 1])
print(paste0("The mean total biomass energy production is ", mean_Total_Biomass_Energy_Production))
```

```
## [1] "The mean total biomass energy production is 286.048930489731"
```

```r
sd_Total_Biomass_Energy_Production <- sd(energy_ts[, 1])
print(paste0("The standard deviation of total biomass energy production is ", sd_Total_Biomass_Energy_Pr
```

```
## [1] "The standard deviation of total biomass energy production is 96.2120922425132"
```

```r
mean_Total_Renewable_Energy_Production <- mean(energy_ts[, 2])
print(paste0("The mean total renewable energy production is ", mean_Total_Renewable_Energy_Production))
```

```
## [1] "The mean total renewable energy production is 409.195206951027"
```

```r
sd_Total_Renewable_Energy_Production <- sd(energy_ts[, 2])
print(paste0("The standard deviation of total renewable energy production is ", sd_Total_Renewable_Energ
```

```
## [1] "The standard deviation of total renewable energy production is 151.422315167649"
```

```r
mean_Hydroelectric_Power_Consumption <- mean(energy_ts[, 3])
print(paste0("The mean hydroelectric power consumption is ", mean_Hydroelectric_Power_Consumption))
```

```
## [1] "The mean hydroelectric power consumption is 79.3568214849921"
```

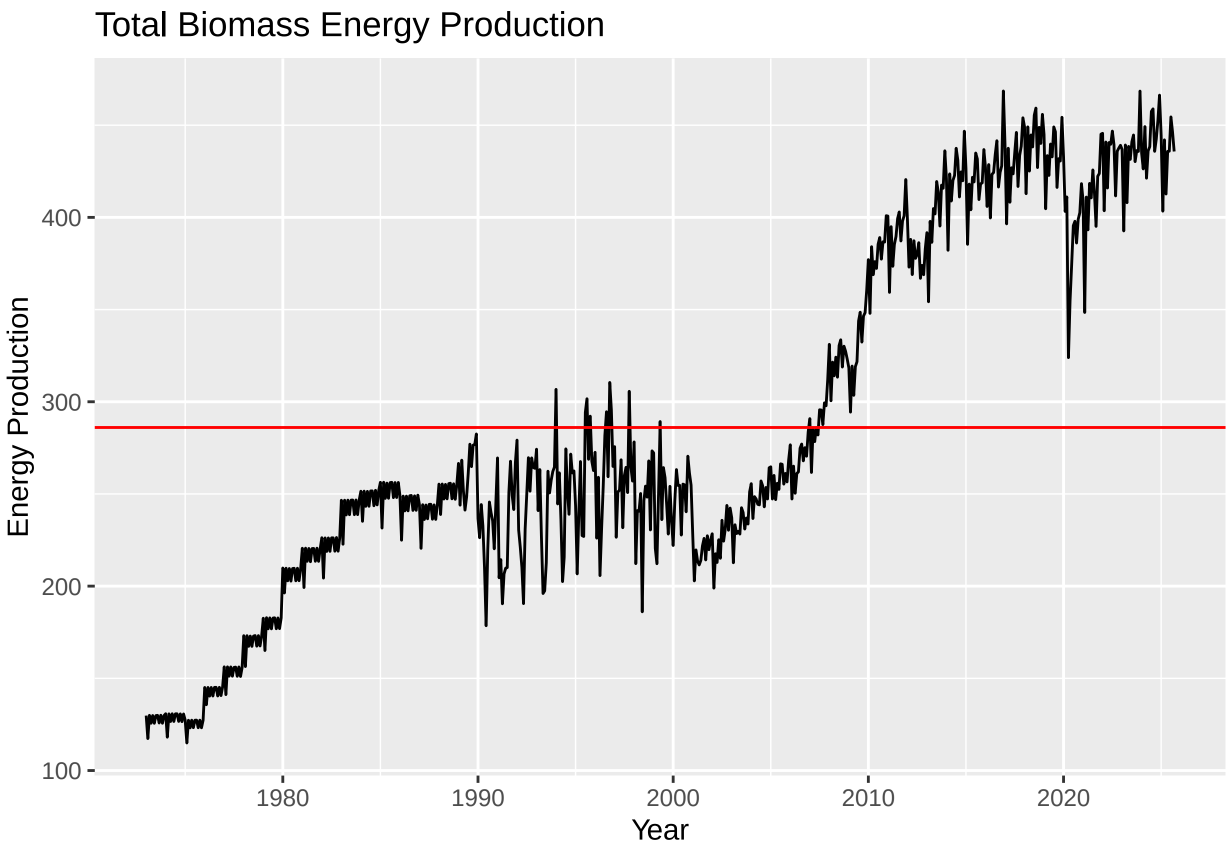
```r
sd_Hydroelectric_Power_Consumption <- sd(energy_ts[, 3])
print(paste0("The standard deviation of hydroelectric power consumption is ", sd_Hydroelectric_Power_Co
```

```
## [1] "The standard deviation of hydroelectric power consumption is 14.1202015109619"
```

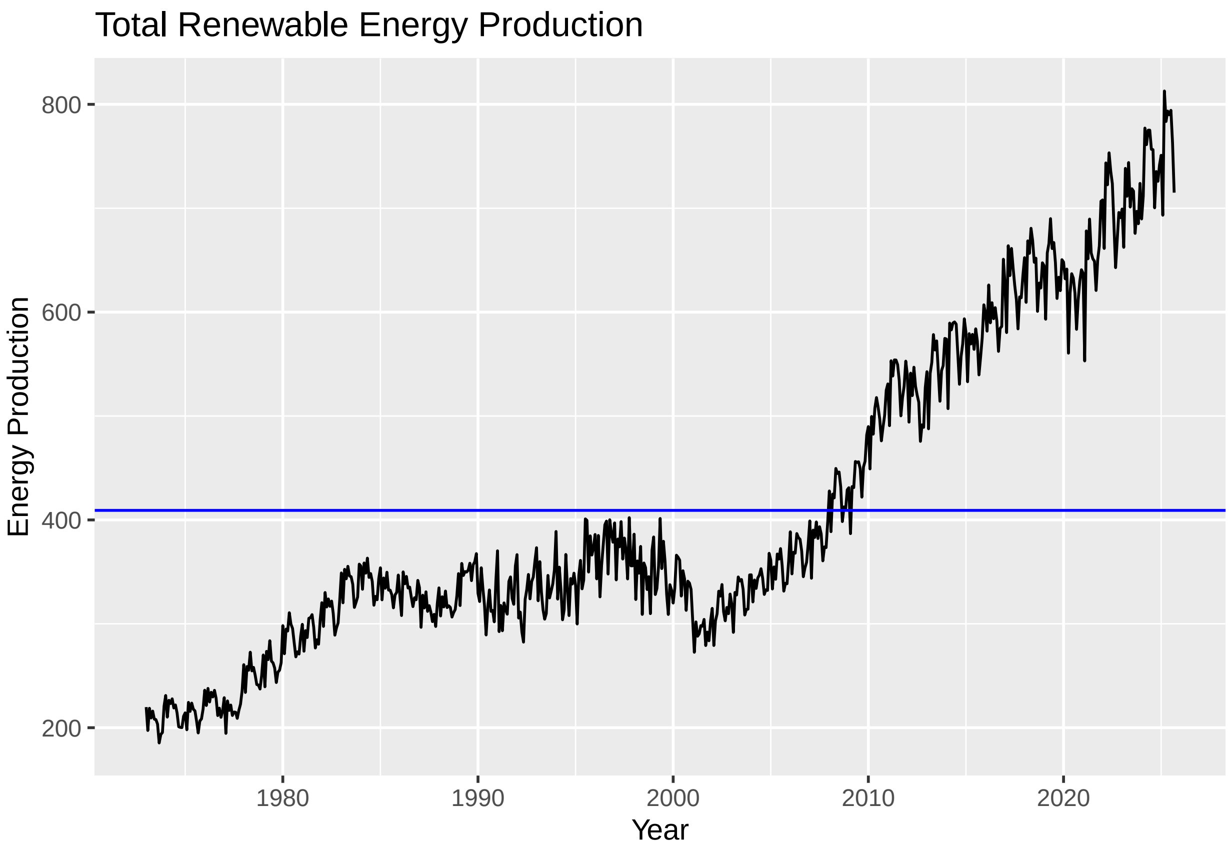
Question 4

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.

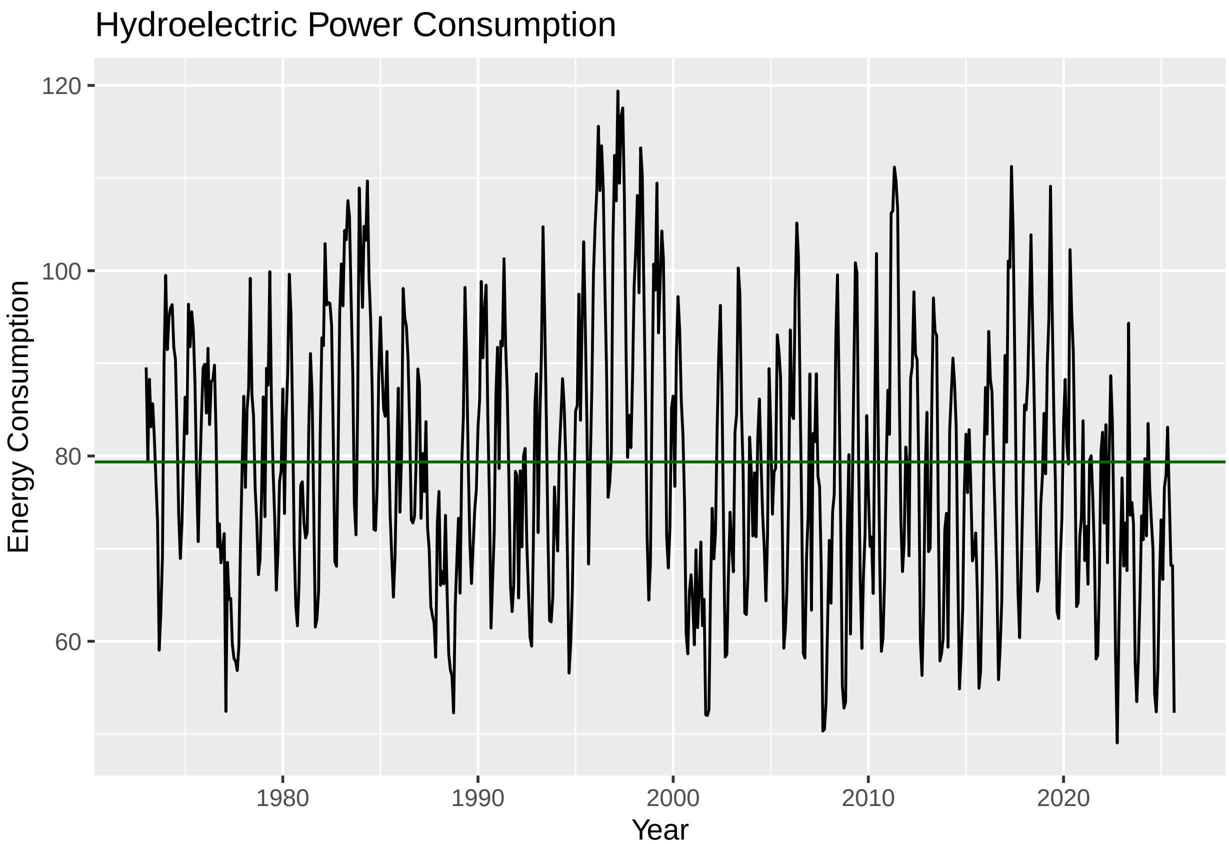
```r
autoplot(energy_ts[, "Total_Biomass_Energy_Production"]) +
  geom_hline(
    yintercept = mean(energy_ts[, "Total_Biomass_Energy_Production"]),
    color = "red"
  ) +
  labs(
    title = "Total Biomass Energy Production",
    x = "Year",
    y = "Energy Production"
  )
```




```r
# 2. Total Renewable Energy Production
autoplot(energy_ts[, "Total_Renewable_Energy_Production"]) +
  geom_hline(
    yintercept = mean(energy_ts[, "Total_Renewable_Energy_Production"]),
    color = "blue",
  ) +
  labs(
    title = "Total Renewable Energy Production",
    x = "Year",
    y = "Energy Production"
  )
```




```r
autoplot(energy_ts[, "Hydroelectric_Power_Consumption"]) +
  geom_hline(
    yintercept = mean(energy_ts[, "Hydroelectric_Power_Consumption"]),
    color = "darkgreen",
  ) +
  labs(
    title = "Hydroelectric Power Consumption",
    x = "Year",
    y = "Energy Consumption"
  )
```



Energy production from biomass seems to have a tight yearly pattern and be steadily increasing from 1973 to about 1988 and once again from about 2001 to about 2020. From about 1988 to about 2001 and after 2020, the yearly pattern seems to be more variable and remains roughly even. Energy production from renewables seems to also have a tight yearly pattern and overall increased year over year modestly from 1972 to about 1984, at which point it stagnates until about 2001, after which energy production increases faster. Energy consumption from hydroelectric power has both a yearly pattern and another pattern with a period of about 7-8 years. From 1973 to the present, on average, energy consumption from hydroelectric power has remained fairly constant.

Question 5

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

```
cor_energy_matrix <- cor(energy_ts)
cor_energy_matrix
```

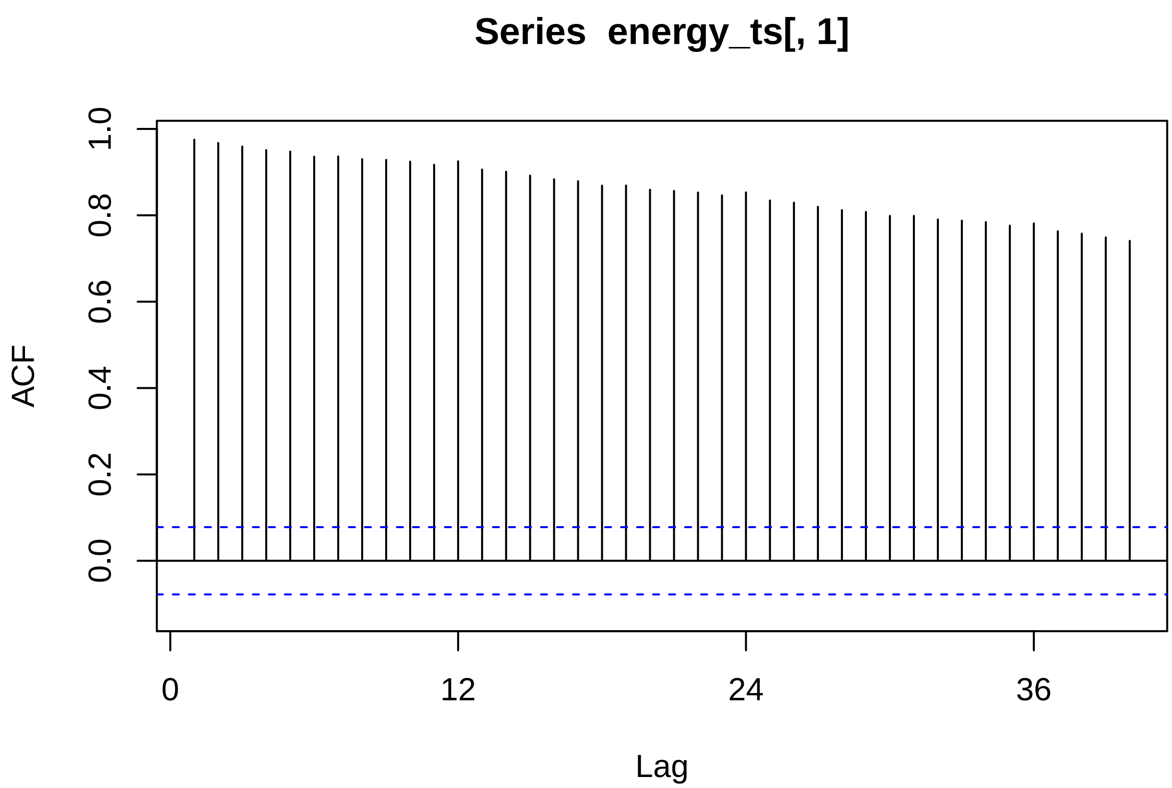
```
##                                   Total_Biomass_Energy_Production
## Total_Biomass_Energy_Production                         1.0000000
## Total_Renewable_Energy_Production                       0.9652985
## Hydroelectric_Power_Consumption                        -0.1347374
##                                   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, while the hydroelectric power consumption does not seem to be correlated to either. The correlation between total biomass energy production and total renewable energy production is 0.965, so strongly positively correlated. The correlation between total biomass energy production and hydroelectric power consumption is - 0.135, so weakly and negatively correlated. The correlation between total renewable energy production and hydroelectric power consumption is -0.058, so essentially uncorrelated.

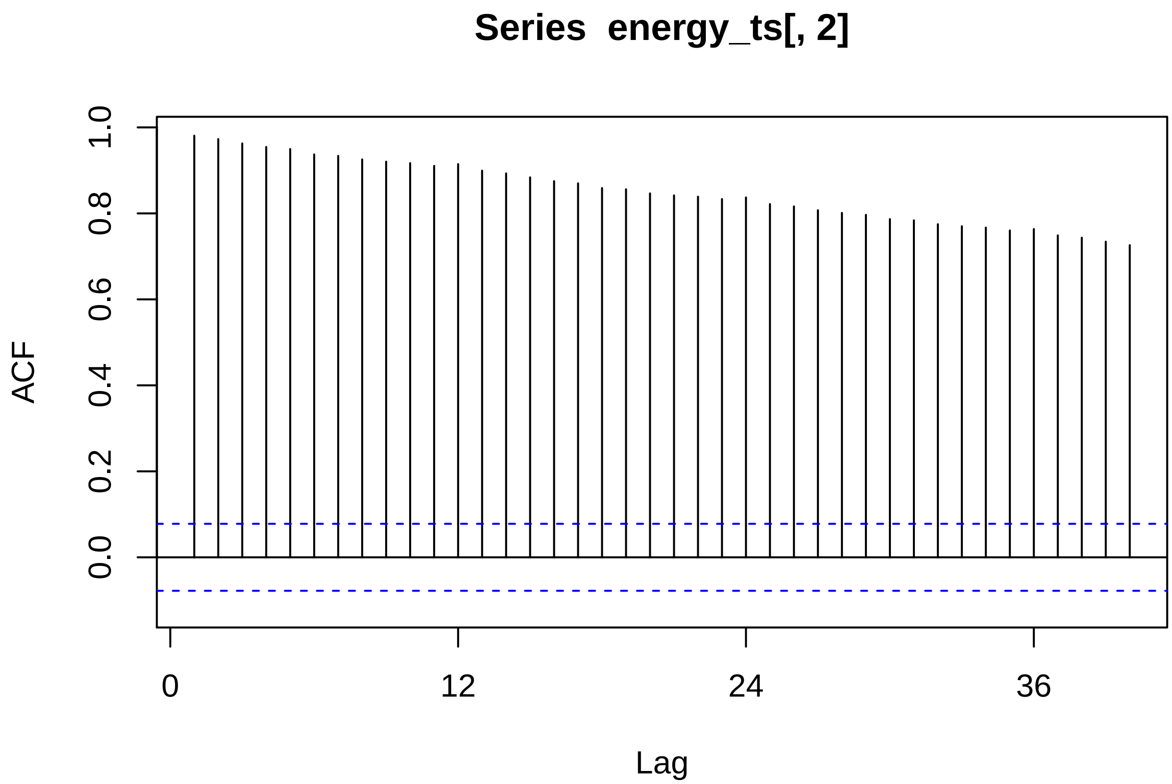
Question 6

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?

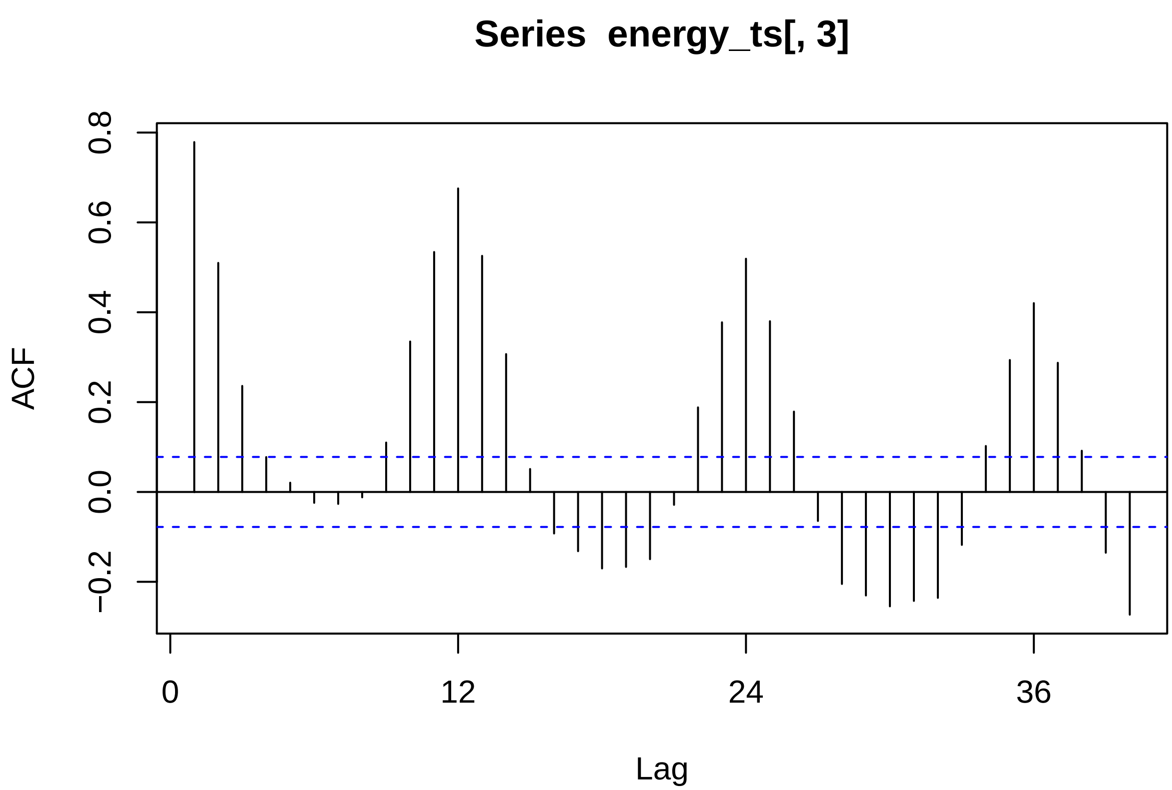
```
acf_Total_Biomass_Energy_Production = Acf(energy_ts[,1], lag.max = 40, plot = TRUE)
```




```
acf_Total_Renewable_Energy_Production = Acf(energy_ts[,2], lag.max = 40, plot = TRUE)
```




```r
acf_Hydroelectric_Power_Consumption = Acf(energy_ts[,3], lag.max = 40, plot = TRUE)
```

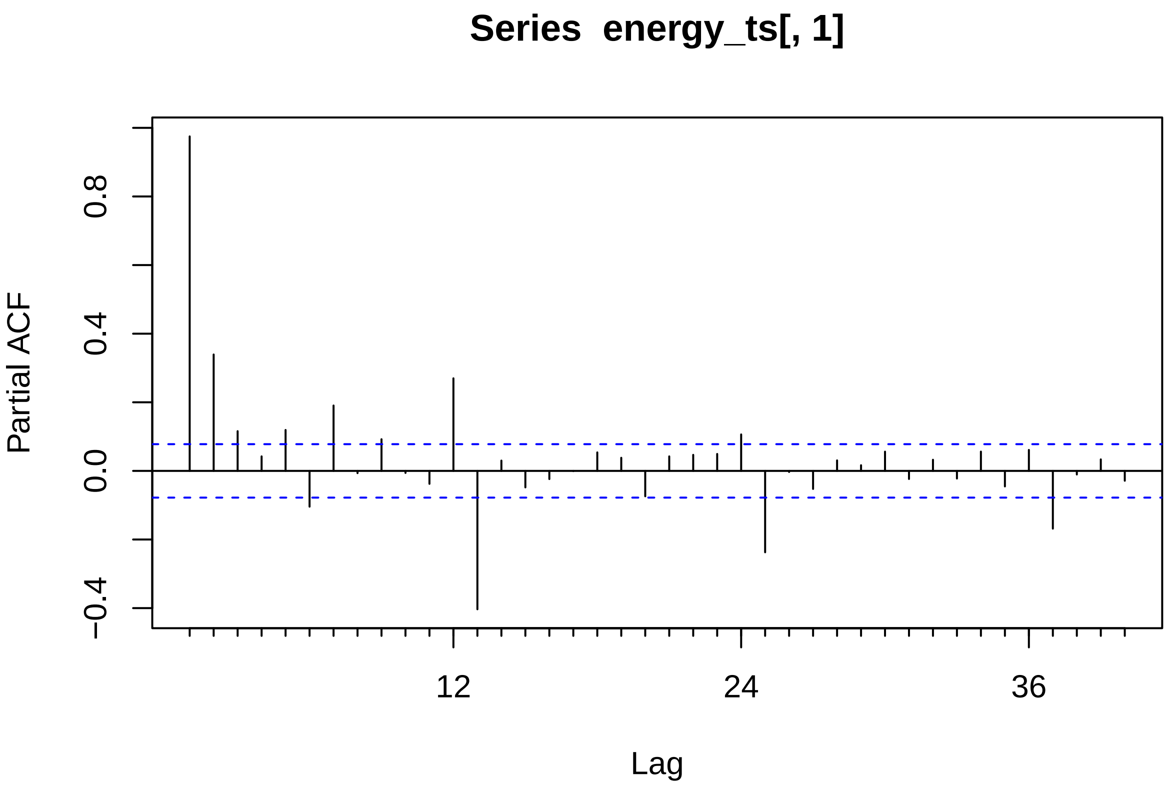


Total biomass energy production and total renewable energy production have roughly the same behavior while the hydroelectric power consumption has very different behavior. The autocorrelation is very high at first and slowly decays to 0.8 for both the total biomass energy production and total renewable energy production. The autocorrelation for the hydroelectric power consumption resembles a sine wave that decreases in mean over time. This means that past values are good predictors of future values for total biomass energy production and total renewable energy production and the trend for hydroelectric power consumption exhibits high seasonality over the course of a year, values exactly a year apart tend to be similar although this relation decays over years.

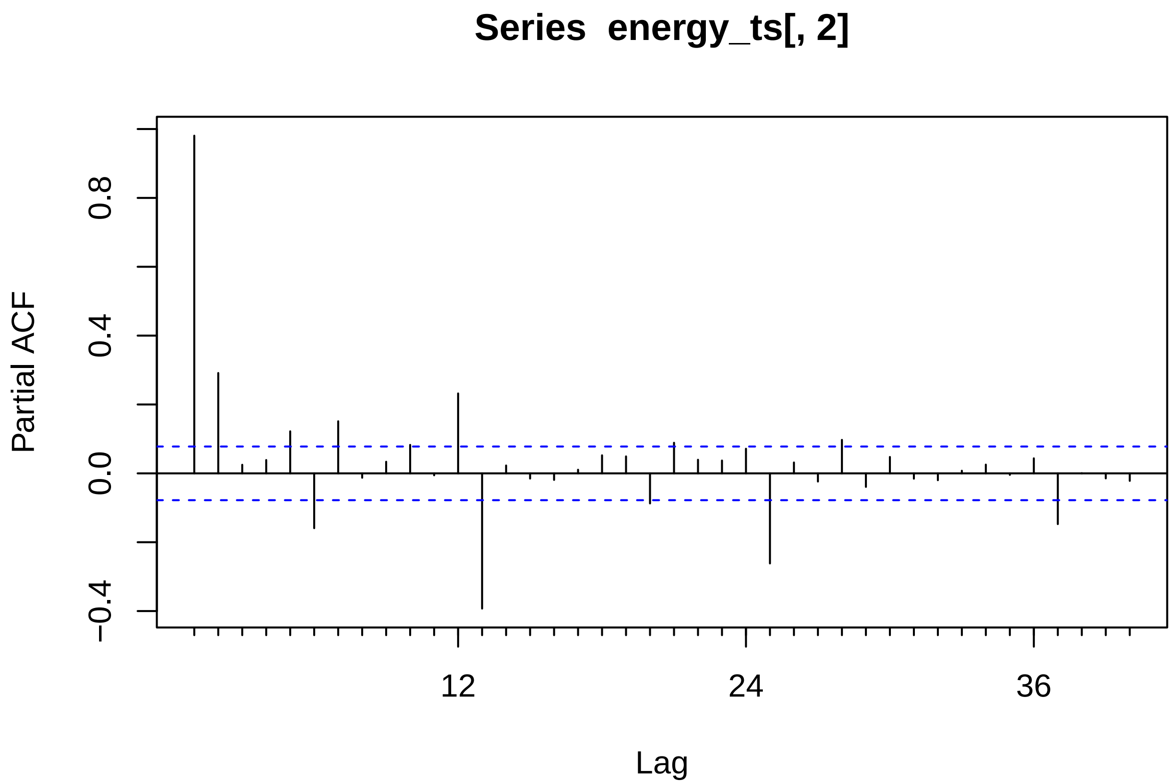
Question 7

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

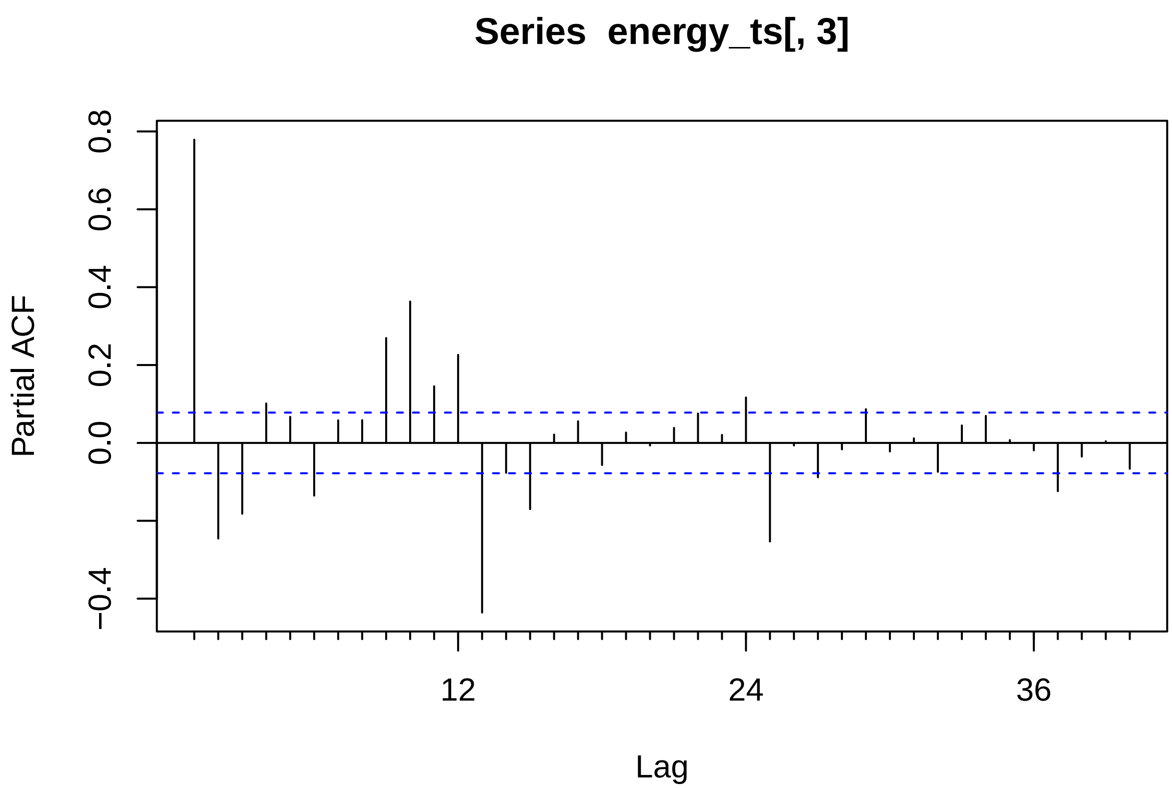
```
pacf_Total_Biomass_Energy_Production = Pacf(energy_ts[,1], lag.max = 40, plot = TRUE)
```




```
pacf_Total_Renewable_Energy_Production = Pacf(energy_ts[,2], lag.max = 40, plot = TRUE)
```




```
pacf_Hydroelectric_Power_Consumption = Pacf(energy_ts[,3], lag.max = 40, plot = TRUE)
```



The partial autocorrelation functions all look essentially the same, starting near 1 and very quickly decaying roughly to a sine wave centered at 0 and with an amplitude that decreases to below 0.1.