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

Assignment 2 - Due date 01/27/26

Aisha

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
#install.packages(c("forecast", "tseries", "dplyr"))
library(forecast)
```

```
## Warning: package 'forecast' was built under R version 4.5.2

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

```r
library(tseries)
```

```
## Warning: package 'tseries' was built under R version 4.5.2
```

```r
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
```

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
#install.packages("readxl")
#install.packages("openxlsx")
library(readxl)
library(openxlsx)
energy_data1 <- read_excel(path="/Users/meilishen/Documents/TimeSeries/TSA_Sp26/Data/Table_10.1_Renewabl
```

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

```
#Now let's extract the column names from row 11
read_col_names <- read_excel(path="/Users/meilishen/Documents/TimeSeries/TSA_Sp26/Data/Table_10.1_Renewa
```

```
## 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>, ...
```

```r
energy_data2 <- read.xlsx(xlsxFile="/Users/meilishen/Documents/TimeSeries/TSA_Sp26/Data/Table_10.1_Rene

read_col_names2  <- read.xlsx(xlsxFile="/Users/meilishen/Documents/TimeSeries/TSA_Sp26/Data/Table_10.1_

#Assign the column names to the data set
colnames(energy_data2) <- read_col_names2

#Visualize the first rows of the data set
head(energy_data2)
```

```
##   Month Wood Energy Production Biofuels Production
## 1 26665                129.630       Not Available
## 2 26696                117.194       Not Available
## 3 26724                129.763       Not Available
## 4 26755                125.462       Not Available
## 5 26785                129.624       Not Available
## 6 26816                125.435       Not Available
##   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 Geothermal Energy Consumption
```

```
## 1                          89.562                          0.490
## 2                          79.544                          0.448
## 3                          88.284                          0.464
## 4                          83.152                          0.542
## 5                          85.643                          0.505
## 6                          82.060                          0.579
##   Solar Energy Consumption Wind Energy Consumption Wood Energy Consumption
## 1            Not Available           Not Available                 129.630
## 2            Not Available           Not Available                 117.194
## 3            Not Available           Not Available                 129.763
## 4            Not Available           Not Available                 125.462
## 5            Not Available           Not Available                 129.624
## 6            Not Available           Not Available                 125.435
##   Waste Energy Consumption Biofuels Consumption
## 1                    0.157        Not Available
## 2                    0.144        Not Available
## 3                    0.176        Not Available
## 4                    0.174        Not Available
## 5                    0.210        Not Available
## 6                    0.176        Not Available
##   Total Biomass Energy Consumption Total Renewable Energy Consumption
## 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
```

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
library(dplyr)
timeseries_df <- energy_data1 [, c( "Total Biomass Energy Production", "Total Renewable Energy Producti
head(timeseries_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().

```
energy_ts <- ts(timeseries_df, start=c(1931,1),frequency=12 )

head(energy_ts[, 1:3])
```

```
##          Total Biomass Energy Production Total Renewable Energy Production
## Jan 1931                         129.787                           219.839
## Feb 1931                         117.338                           197.330
## Mar 1931                         129.938                           218.686
## Apr 1931                         125.636                           209.330
## May 1931                         129.834                           215.982
## Jun 1931                         125.611                           208.249
##          Hydroelectric Power Consumption
## Jan 1931                          89.562
## Feb 1931                          79.544
## Mar 1931                          88.284
## Apr 1931                          83.152
## May 1931                          85.643
## Jun 1931                          82.060
```

```
dim(energy_ts)
```

```
## [1] 633   3
```

Question 3

Compute mean and standard deviation for these three series.

```
#the number in the apply function refers to 1 for row and 2 for column
mean_series<-apply(energy_ts, 2, mean)
sd_series<-apply(energy_ts, 2, sd)

mean_series
```

```
##   Total Biomass Energy Production Total Renewable Energy Production
##                         286.04893                         409.19521
##   Hydroelectric Power Consumption
##                          79.35682
```

```
sd_series
```

```
##   Total Biomass Energy Production Total Renewable Energy Production
##                          96.21209                         151.42232
##   Hydroelectric Power Consumption
##                          14.12020
```

Question 4

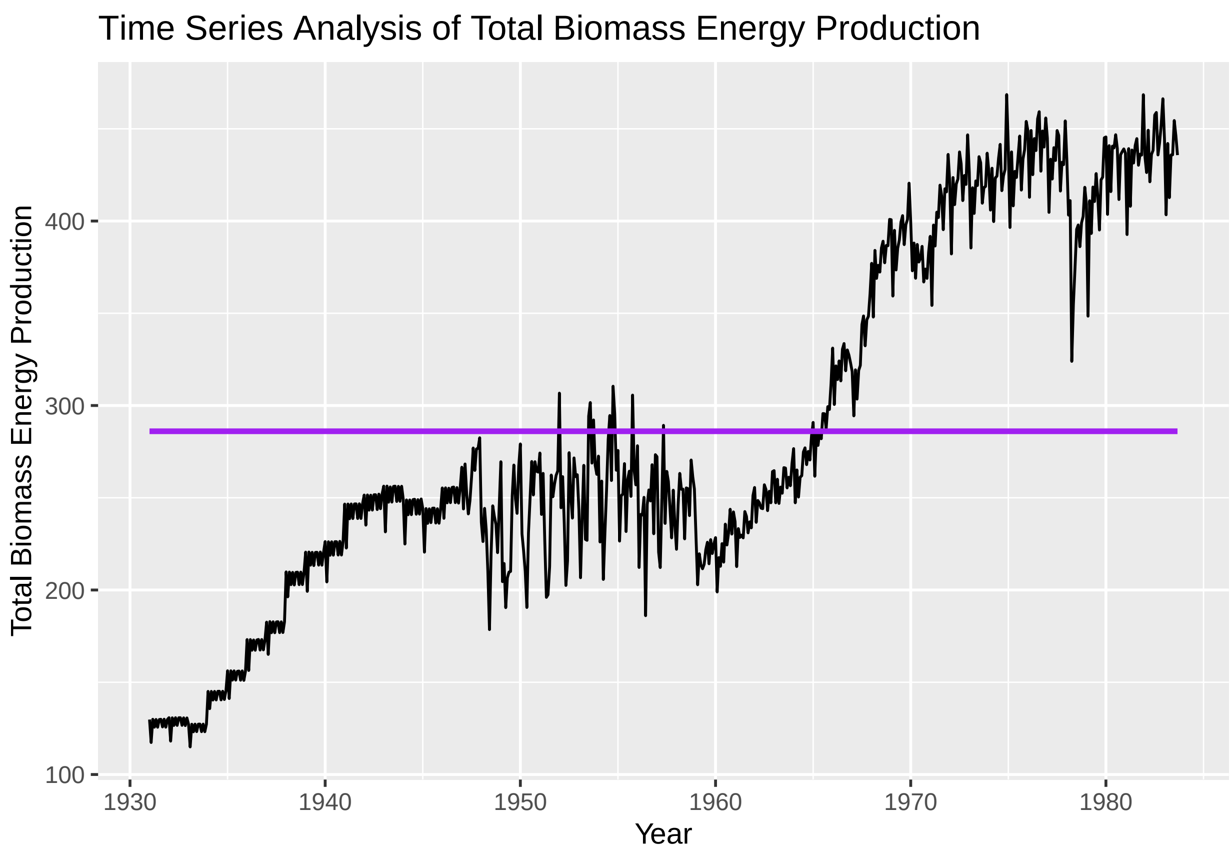
Display and interpret the time series plot for each of these variables. Try to make your plot as informative as possible by writing titles, labels, etc. For each plot add a horizontal line at the mean of each series

```
#using package ggplot2
library(ggplot2)
```

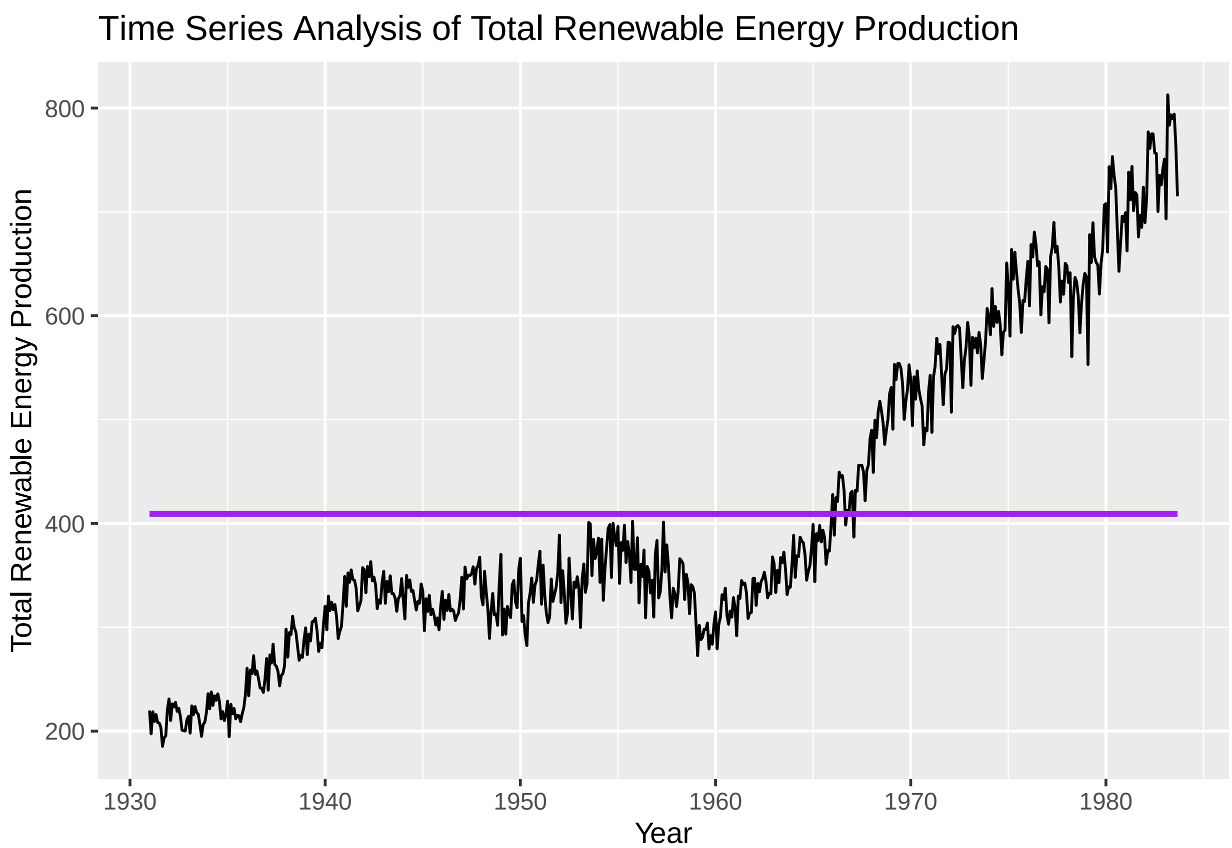
```
## Warning: package 'ggplot2' was built under R version 4.5.2
```

```
autoplot(energy_ts[,1])+labs(title = "Time Series Analysis of Total Biomass Energy Production",
       x = "Year",
       y = "Total Biomass Energy Production") + geom_line(aes(y = mean_series[1]), color = "purple", si
```

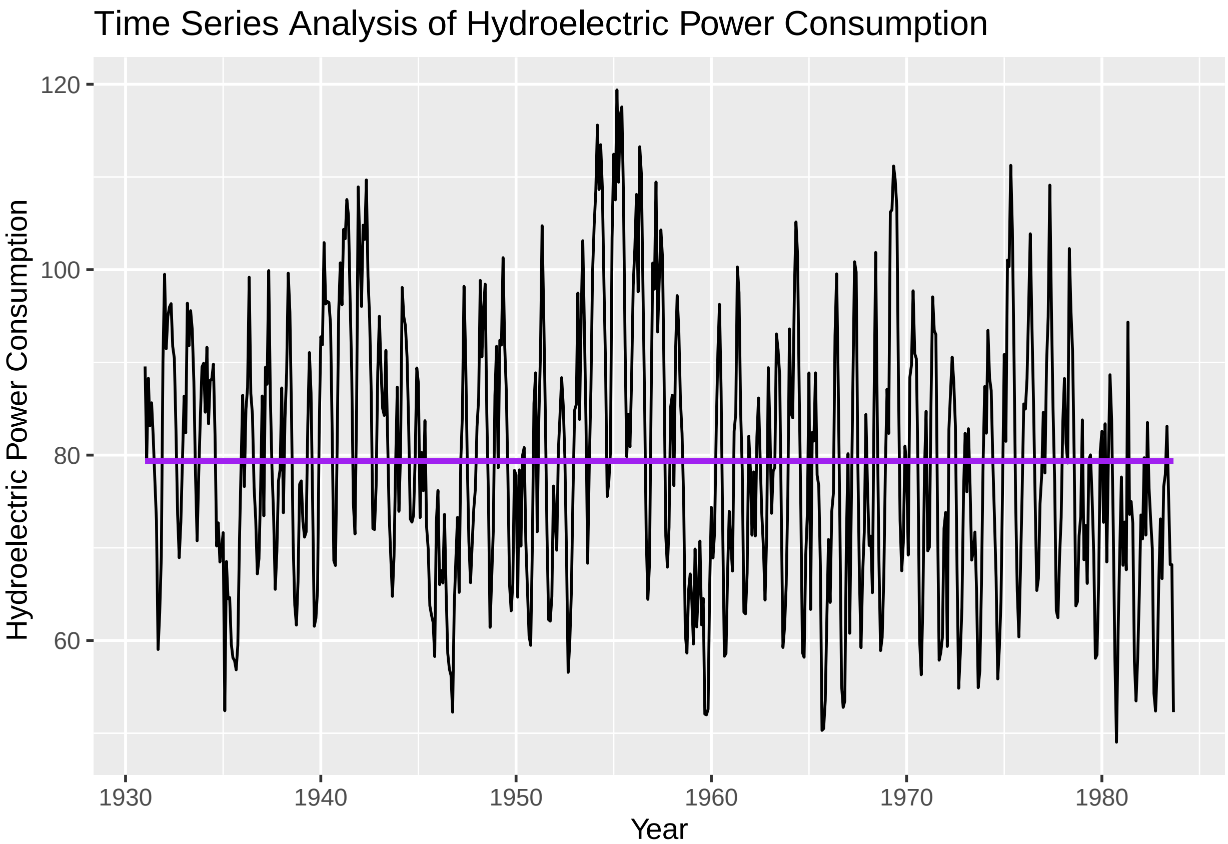
```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once per session.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```




```
autoplot(energy_ts[,2])+labs(title = "Time Series Analysis of Total Renewable Energy Production",
       x = "Year",
       y = "Total Renewable Energy Production") +geom_line(aes(y = mean_series[2]), color = "purple", si
```




```
autoplot(energy_ts[,3])+labs(title = "Time Series Analysis of Hydroelectric Power Consumption ",
       x = "Year",
       y = "Hydroelectric Power Consumption ") +geom_line(aes(y = mean_series[3]), color = "purple", si
```



Question 5

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

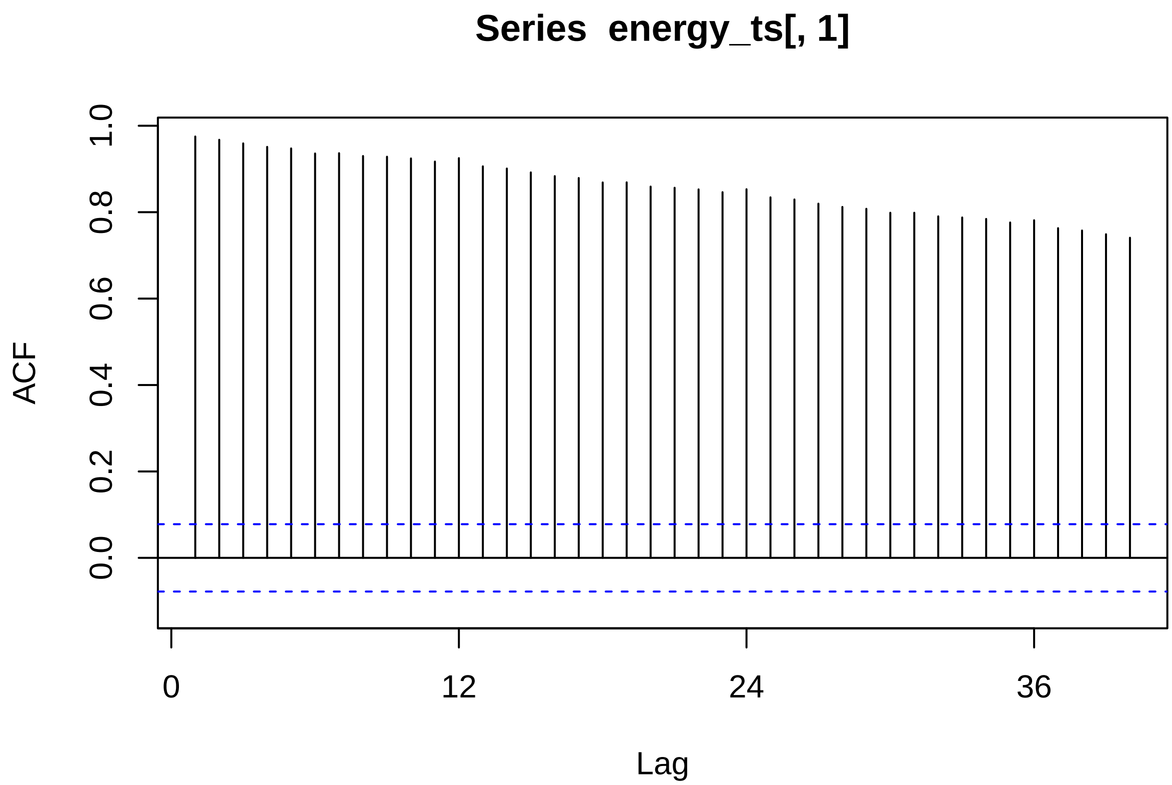
```
cor(energy_ts) #using a linear correlation function, we see that biomass is highly correlated with rene
```

```
##                                     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
```

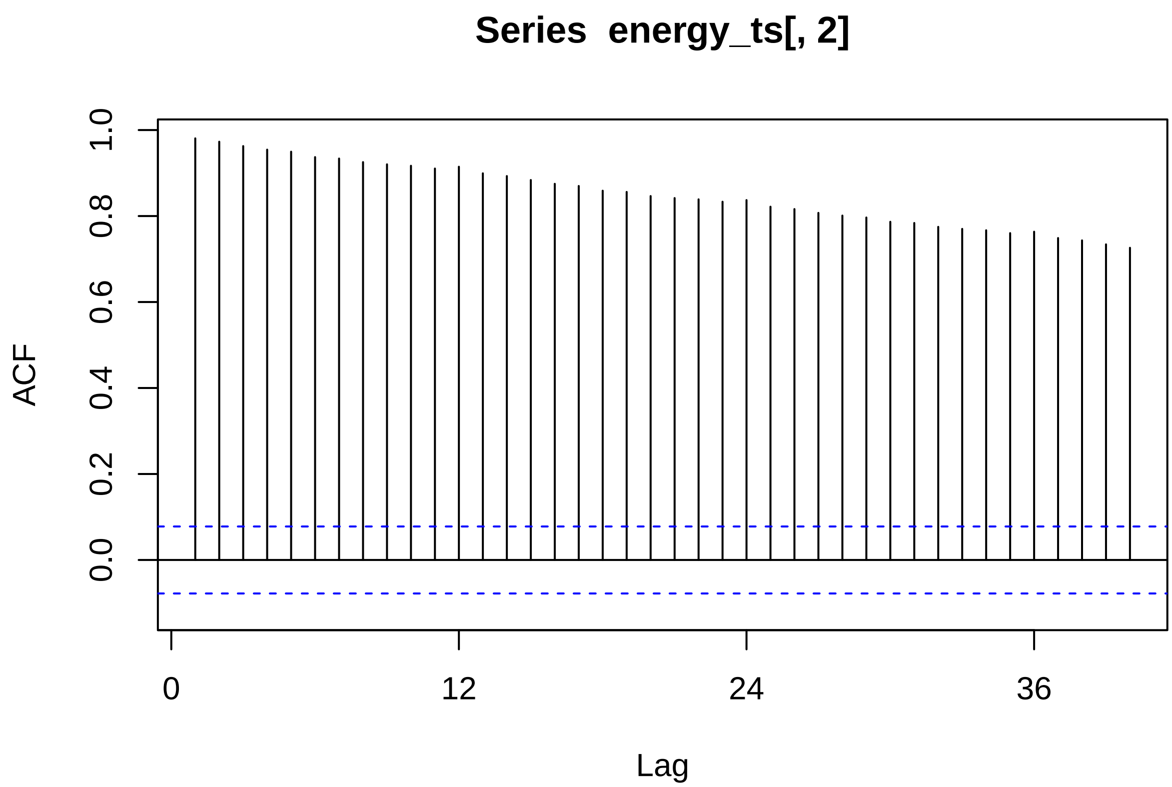
Question 6

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?

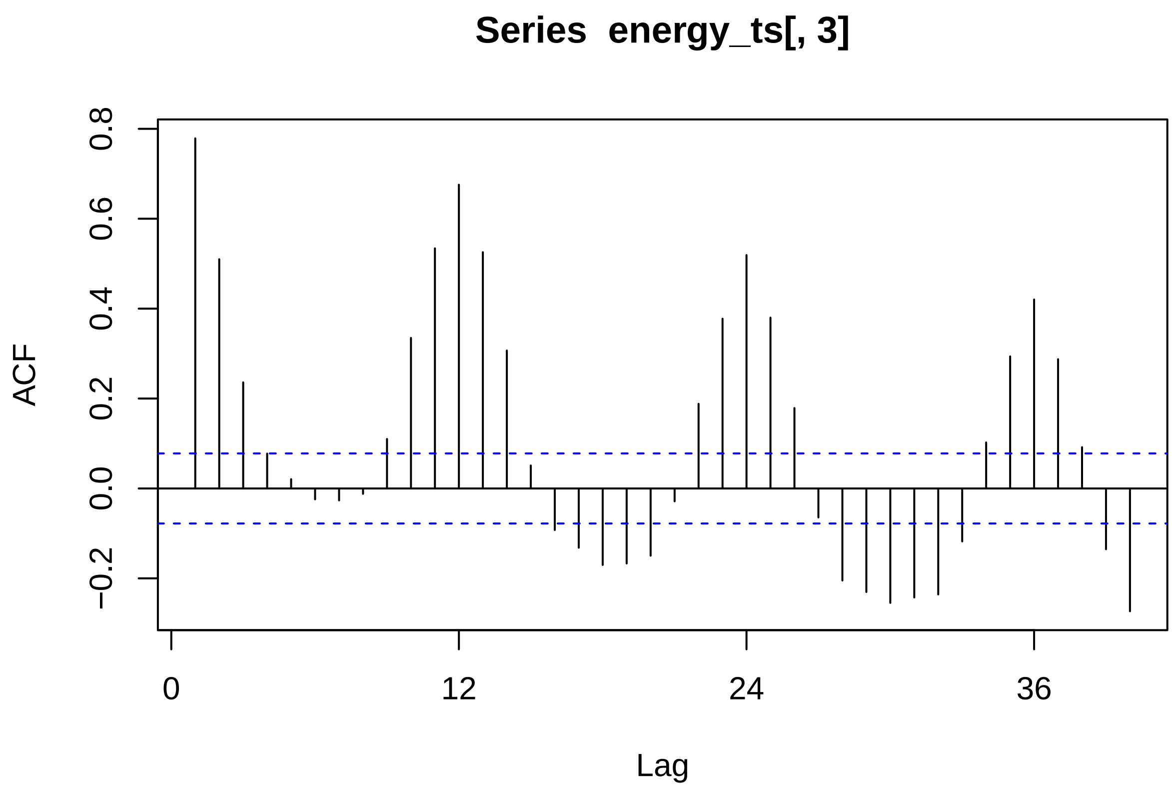
```r
library(forecast)
Acf(energy_ts[,1],lag.max=40)
```




```r
Acf(energy_ts[,2],lag.max=40)
```




```
Acf(energy_ts[,3],lag.max=40)
```

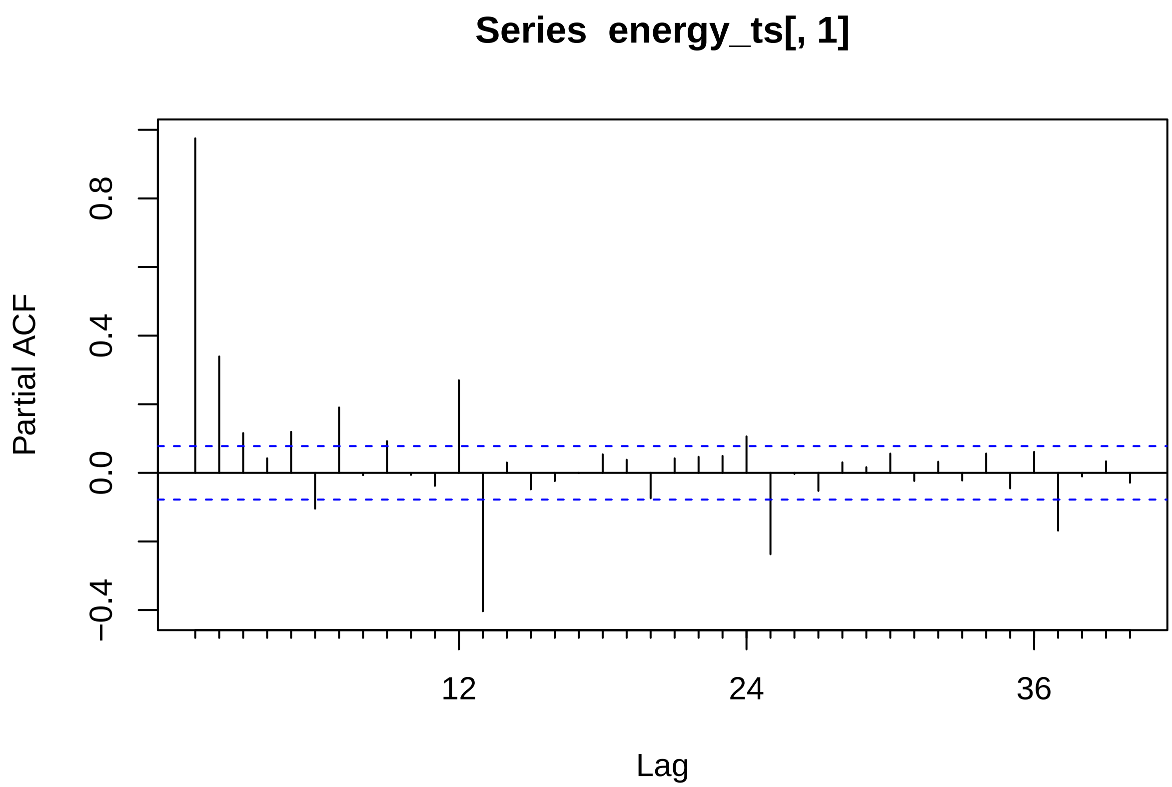


ACF of both Total Biomass Energy Production and Total Renewable Energy Production shows a gradual decay of positive autocorrelations, meaning there is a strong deterministic trend and it is not just random noise. All lags are significant as they are above the significant bound. The ACF of Hydroelectric Power Consumption shows persistent autocorrelation across many lags with spikes at the lag 12, 24, and 36 that indicate that there is a seasonality.

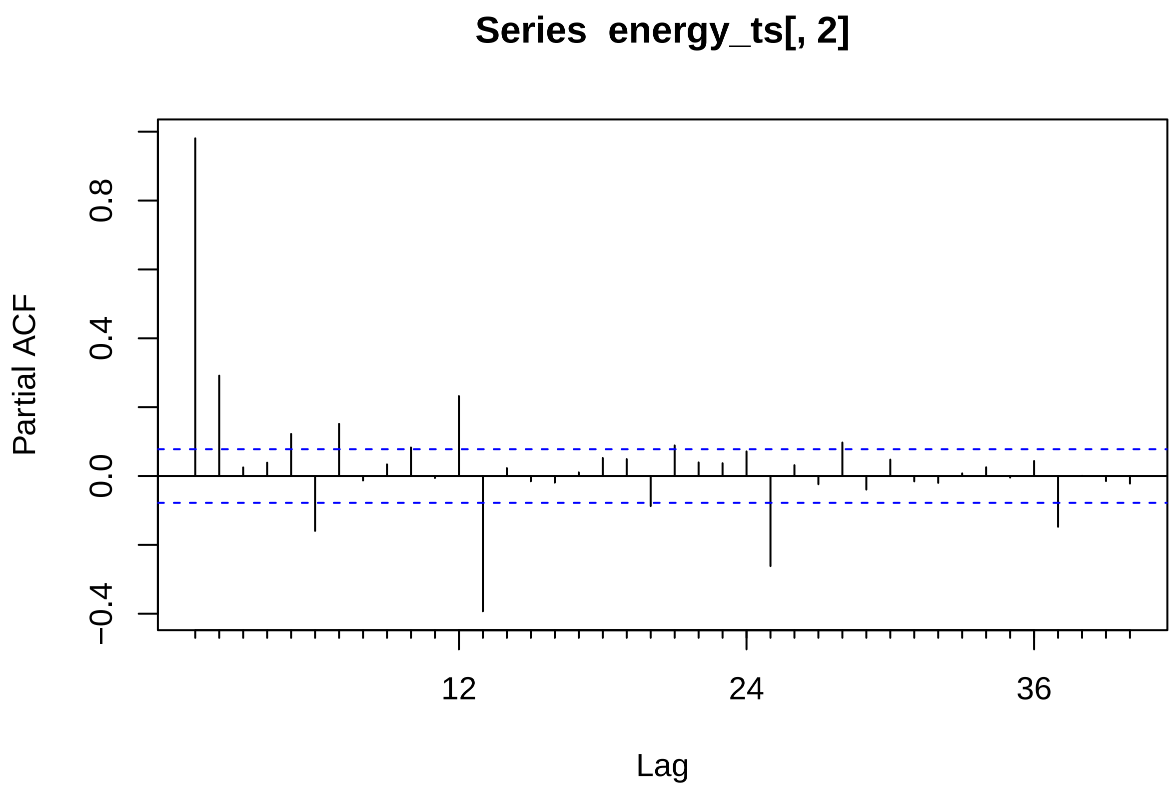
Question 7

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

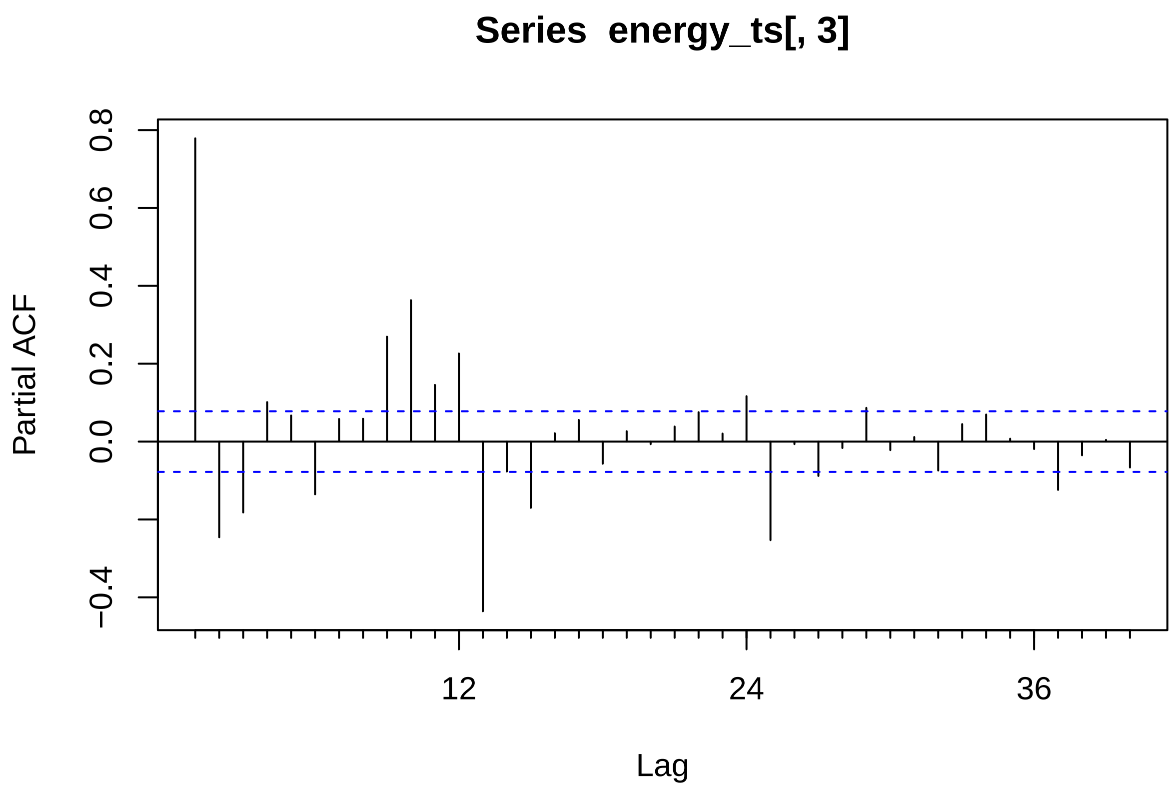
```
Pacf(energy_ts[,1],40)
```




```
Pacf(energy_ts[,2],40)
```




```
Pacf(energy_ts[,3],40)
```



PACF of Total Biomass Energy Production shows that only lags 1-3, 4, 6, and 23 are positively correlated. Both lags 13 and 25 negatively correlated, and this indicate the seasonal autoregressive effects. The PACF of Total Renewable Energy Production exhibits a pattern similar to that of biomass energy production. Significant positive partial autocorrelations are observed at low lags (up to lag 2) and at selected higher lags (5, 7, and 12), suggesting short-term autoregressive behavior combined with seasonal effects. Strong negative partial autocorrelations at lags 6, 13, 25, and 37 further reinforce the presence of seasonal autoregressive structure, with recurring dependence at approximately annual intervals. The PACF of Hydroelectric Power Consumption shows strong positive cpartial autocorrelation at lag 1 and higher lags 9,10, and 24 -> short-term persistence and significant negative correlation at lags 2,3,6,13,15,25,and 37. Compared to ACF, these PACFs indicate short-term and seasonal autoregressive dependence across all three series, while the slowly decaying ACFs with strong seasonal spikes indicates non-stationary and annual seasonality.