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

Assignment 2 - Due date 01/27/26

Lauren Xu

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(readxl)
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)
```

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
getwd()
```

```
## [1] "/Users/xxx/R/Time Series_R/TSA/Assignments"
```

```
energy_data1 <- read_excel(path="/Users/xxx/R/Time Series_R/TSA/Data/Table_10.1_Renewable_Energy_Product
```

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

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.

```
read_col_names <- read_excel(path="/Users/xxx/R/Time Series_R/TSA/Data/Table_10.1_Renewable_Energy_Produ
```

```
## New names:
## * `` -> `...1`
## * `` -> `...2`
## * `` -> `...3`
## * `` -> `...4`
## * `` -> `...5`
## * `` -> `...6`
## * `` -> `...7`
## * `` -> `...8`
## * `` -> `...9`
```

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

```r
colnames(energy_data1) <- read_col_names

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
ts_energy_data1 <- energy_data1[,4:6]
```

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().

Question 3

Compute mean and standard deviation for these three series.

```r
ts.mean <- colMeans(ts_energy_data1, na.rm = TRUE)
ts.sd <- apply(ts_energy_data1, 2, sd, na.rm = TRUE)

ts.mean
```

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

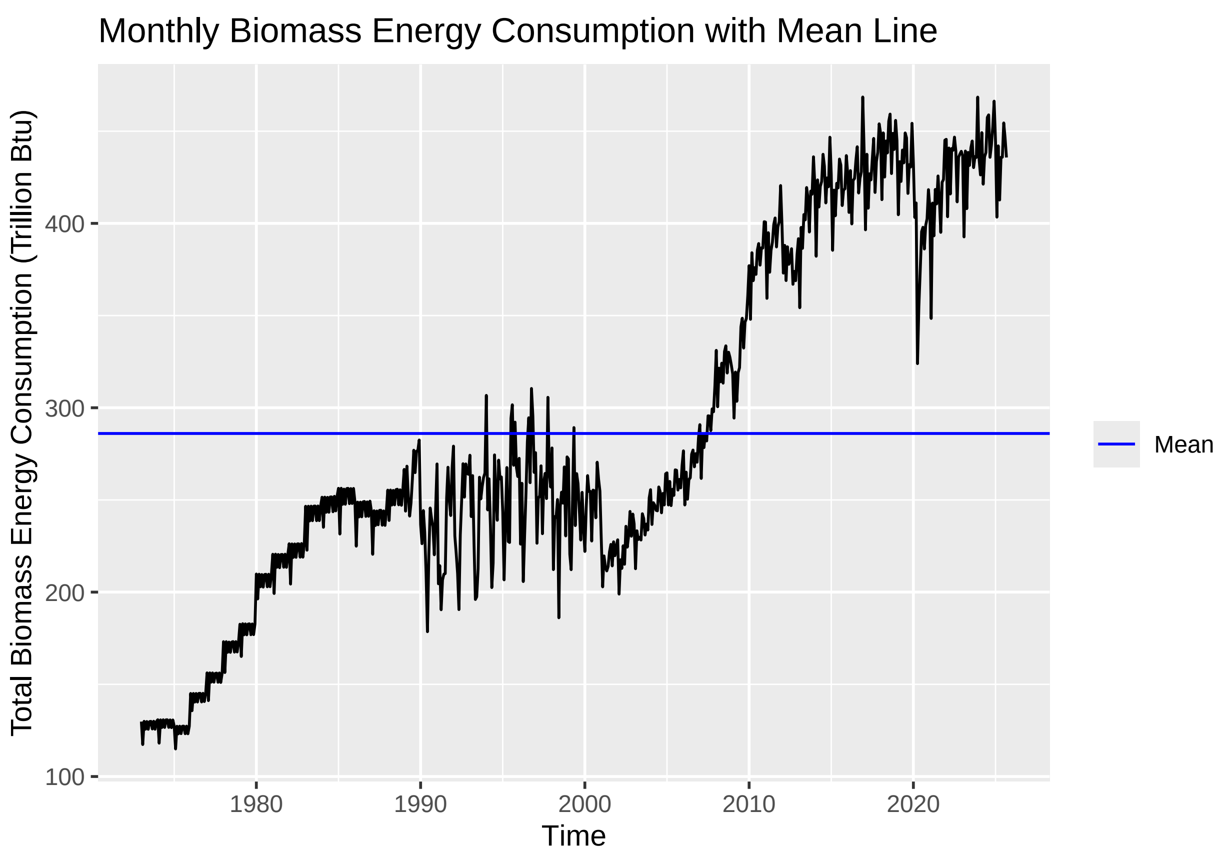
```r
ts.sd
```

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

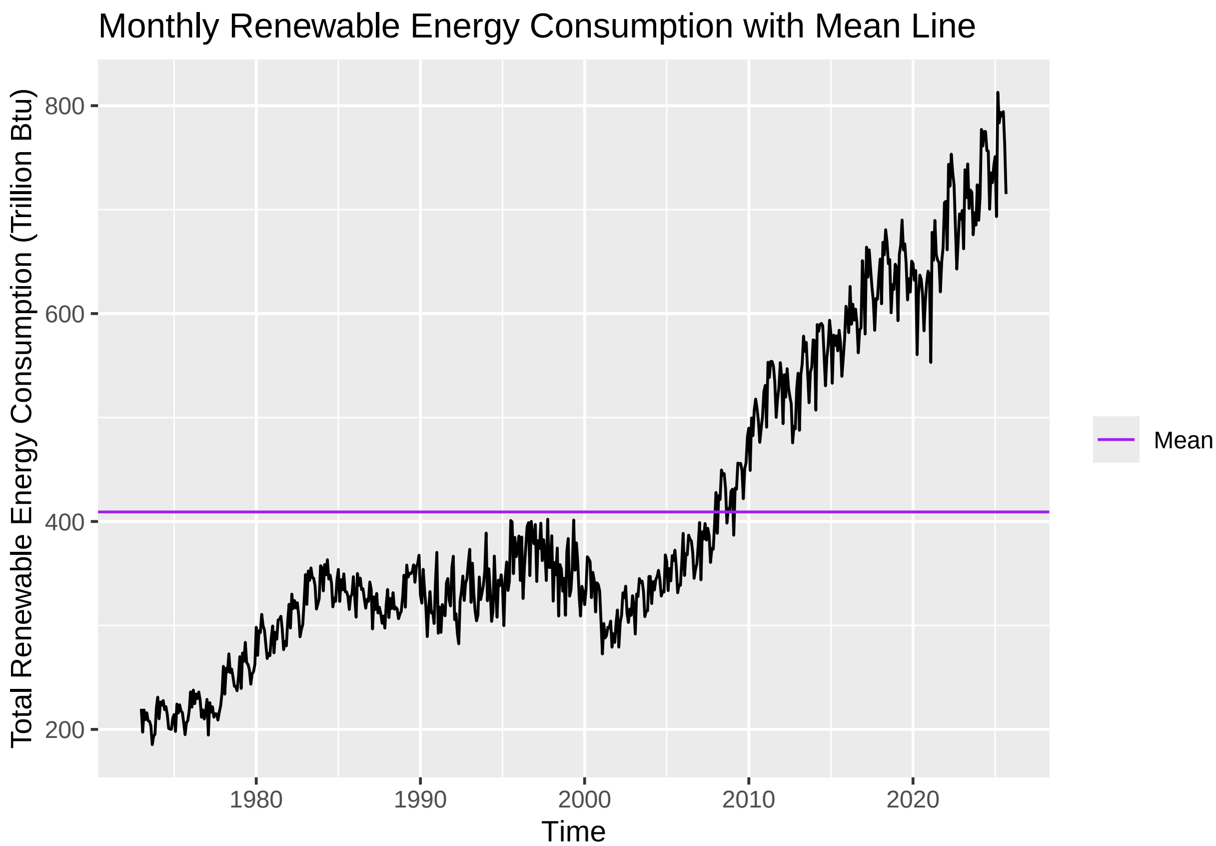
Question 4

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 in a different color.

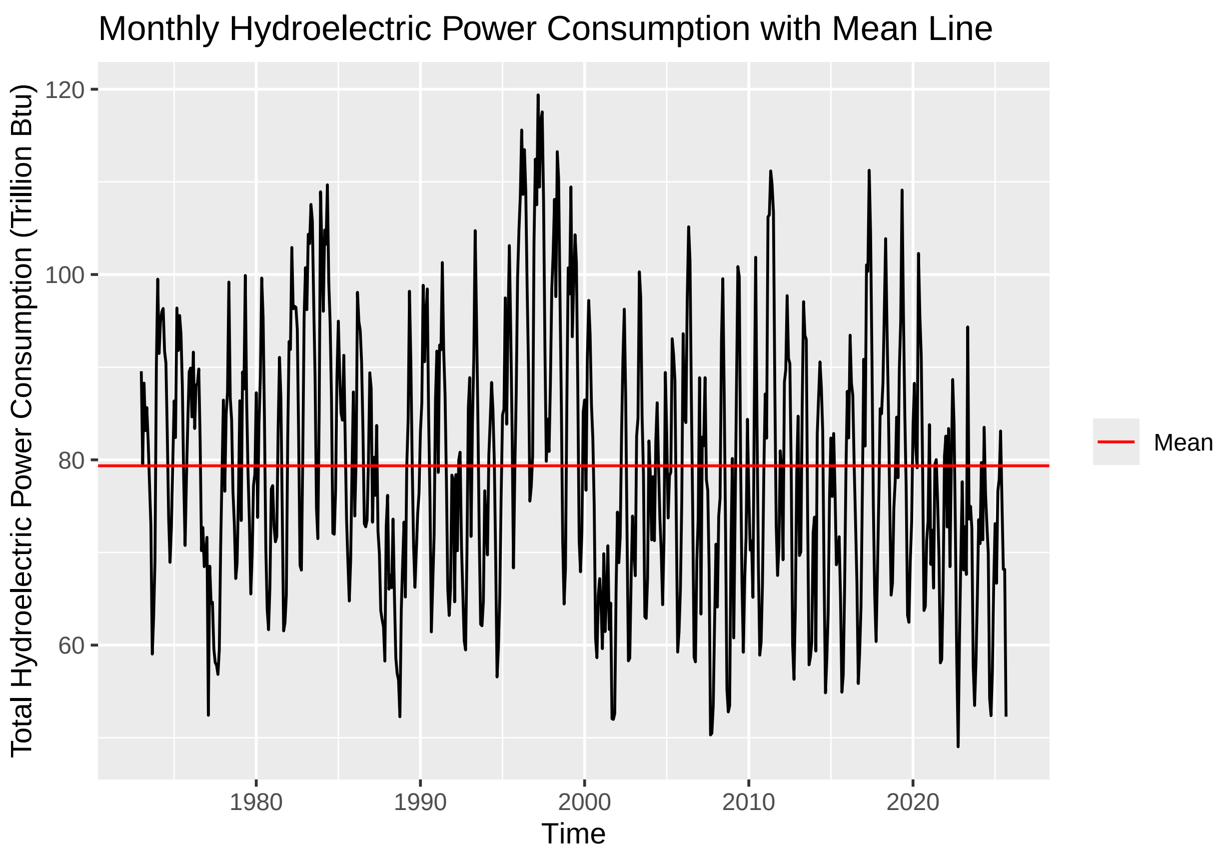
```r
mean_biomass <- mean(ts_energy_data1[,1], na.rm = TRUE)
mean_renewable <- mean(ts_energy_data1[,2], na.rm = TRUE)
mean_hydroelectric <- mean(ts_energy_data1[,3], na.rm = TRUE)


autoplot(ts_energy_data1[,1]) +
  geom_hline(
    aes(yintercept = mean_biomass, color = "Mean")
  )+
  scale_color_manual(values = c("Mean" = "blue"), name = "") +
  xlab("Time") +
  ylab("Total Biomass Energy Consumption (Trillion Btu)") +
  ggtitle("Monthly Biomass Energy Consumption with Mean Line")
```



```r
autoplot(ts_energy_data1[,2]) +
  geom_hline(
  aes(yintercept = mean_renewable, color = "Mean")
  )+
  scale_color_manual(values = c("Mean" = "purple"), name = "") +
  xlab("Time") +
  ylab("Total Renewable Energy Consumption (Trillion Btu)") +
  ggtitle("Monthly Renewable Energy Consumption with Mean Line")
```




```r
autoplot(ts_energy_data1[,3]) +
  geom_hline(
    aes(yintercept = mean_hydroelectric, color = "Mean")
  )+
  scale_color_manual(values = c("Mean" = "red"), name = "") +
  xlab("Time") +
  ylab("Total Hydroelectric Power Consumption (Trillion Btu)") +
  ggtitle("Monthly Hydroelectric Power Consumption with Mean Line")
```



Question 5

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

```
cor.test(ts_energy_data1[,1],
         ts_energy_data1[,2])
```

```
##
##  Pearson's product-moment correlation
##
## data:  ts_energy_data1[, 1] and ts_energy_data1[, 2]
## t = 92.851, df = 631, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.9595516 0.9702413
## sample estimates:
##       cor
## 0.9652985
```

```
cor.test(ts_energy_data1[,1],
         ts_energy_data1[,3])
```

```
##
##  Pearson's product-moment correlation
```

```
##
## data:  ts_energy_data1[, 1] and ts_energy_data1[, 3]
## t = -3.4157, df = 631, p-value = 0.000677
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.21045616 -0.05741173
## sample estimates:
##        cor
## -0.1347374
```

```
cor.test(ts_energy_data1[,2],
         ts_energy_data1[,3])
```

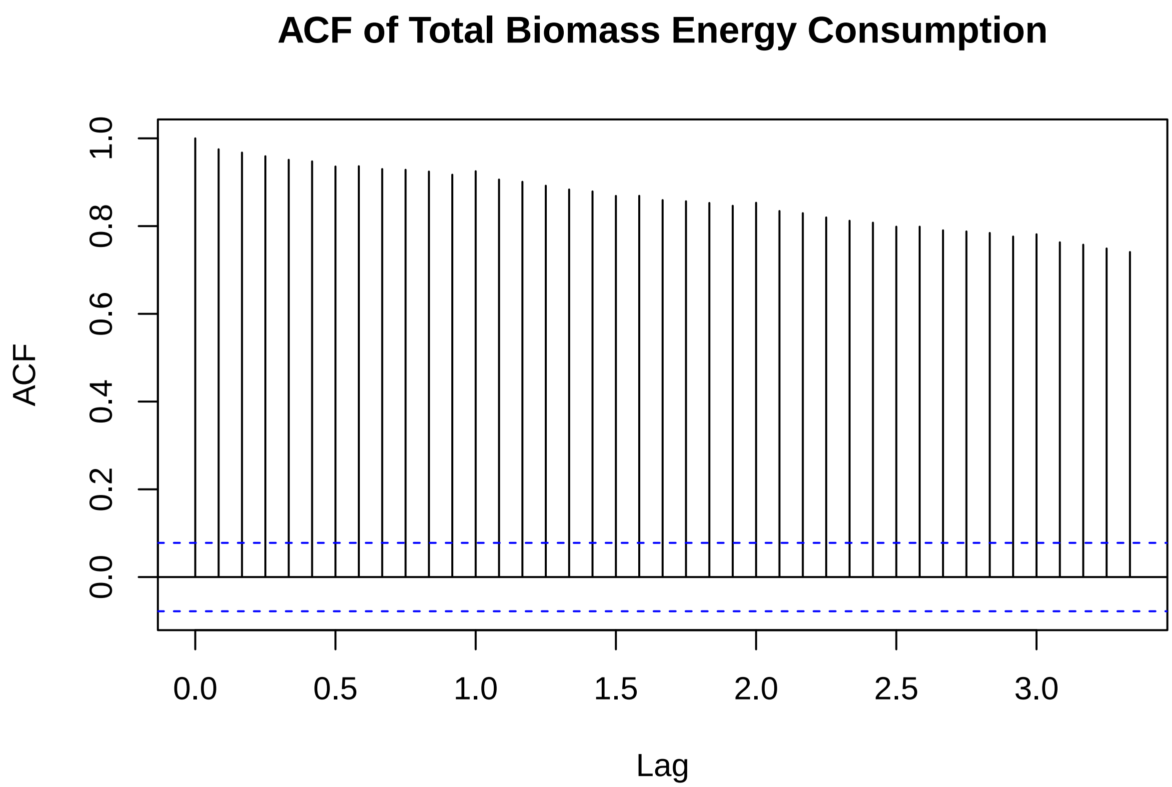
```
##
##  Pearson's product-moment correlation
##
## data:  ts_energy_data1[, 2] and ts_energy_data1[, 3]
## t = -1.4701, df = 631, p-value = 0.142
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.13573488  0.01959335
## sample estimates:
##         cor
## -0.05842436
```

1. Renewable energy and Biomass energy are strongly and significantly correlated, since correlation is 0.965 and p-value<0.05.
2. Biomass is negatively and not strongly correlated with the Hydroelectric power, as the correlations are -0.135. The relationship is statistically significant (p-value<0.05).
3. Renewable energy is also negatively and not strongly correlated (-0.058) with hydroelectric power and the relationship is not significant (p-value>0.05).

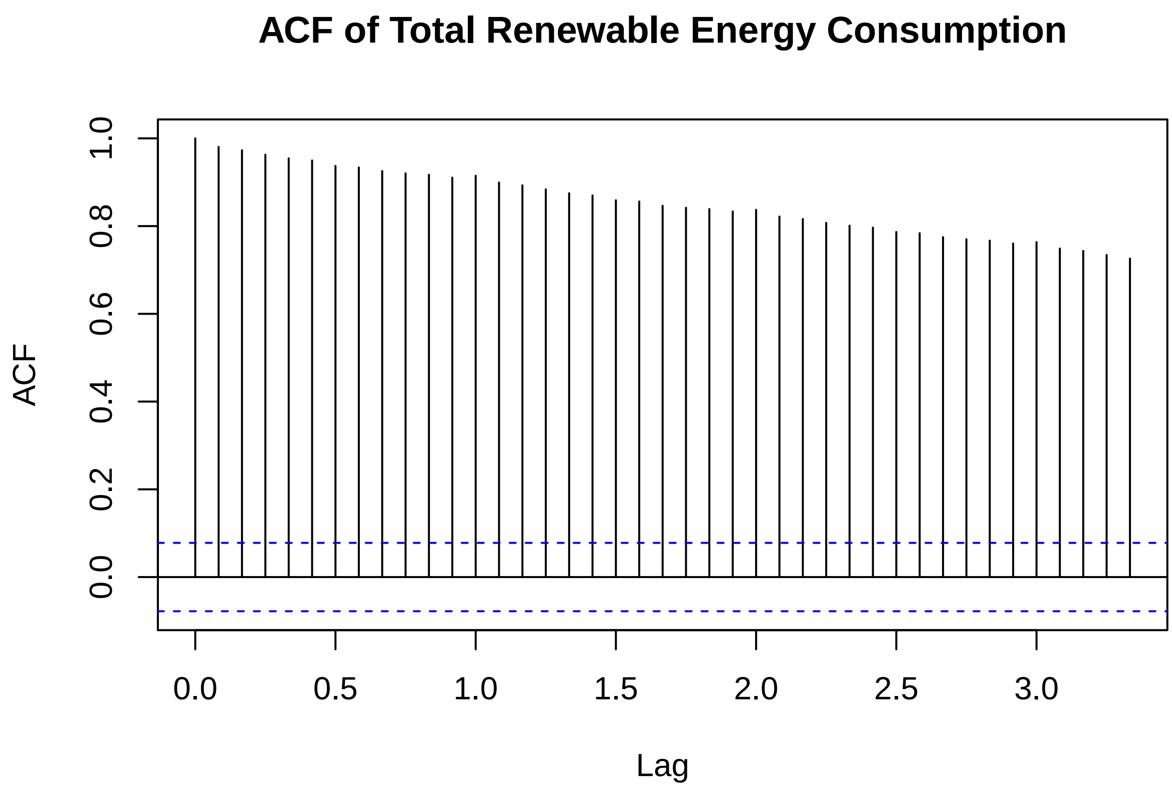
Question 6

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?

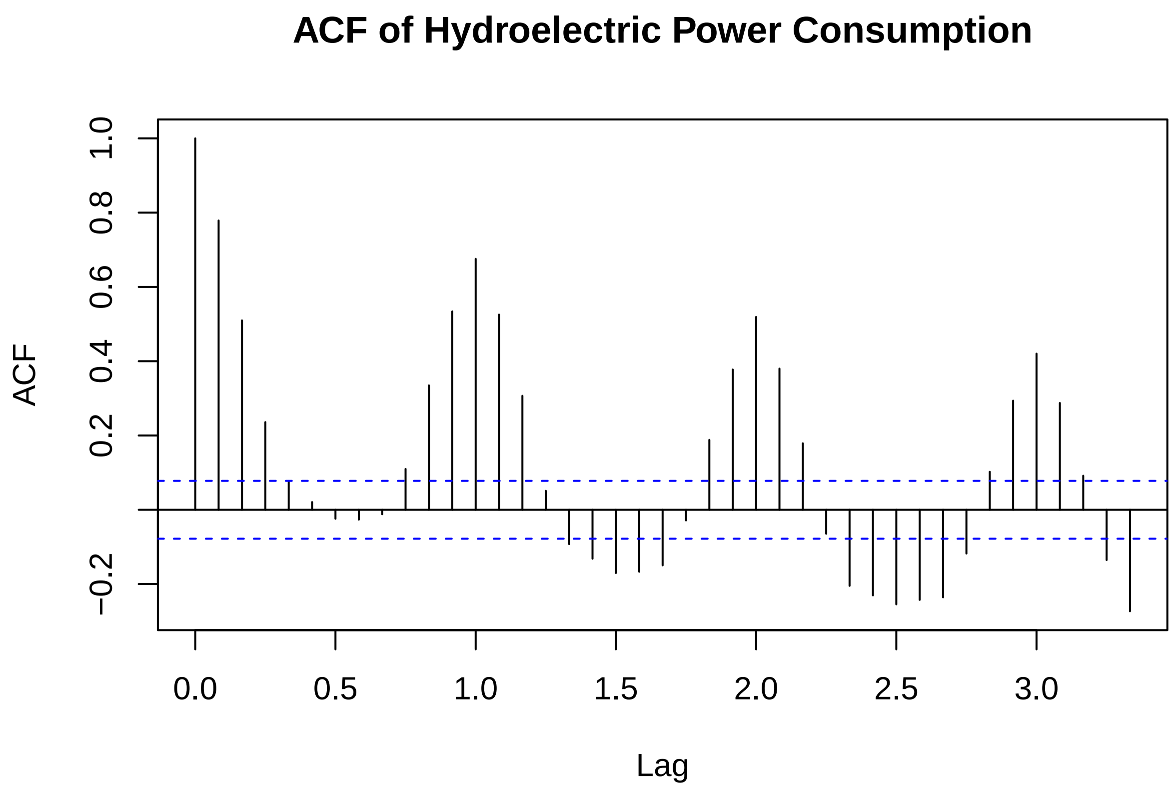
```
acf_biomass <- acf(ts_energy_data1[,1], lag.max = 40, plot = TRUE, main = ("ACF of Total Biomass Energy
```




```
acf_renewable <- acf(ts_energy_data1[,2], lag.max = 40, plot = TRUE, main = ("ACF of Total Renewable En
```




```
acf_hydroelectric <- acf(ts_energy_data1[,3], lag.max = 40, plot = TRUE, main = ("ACF of Hydroelectric P
```



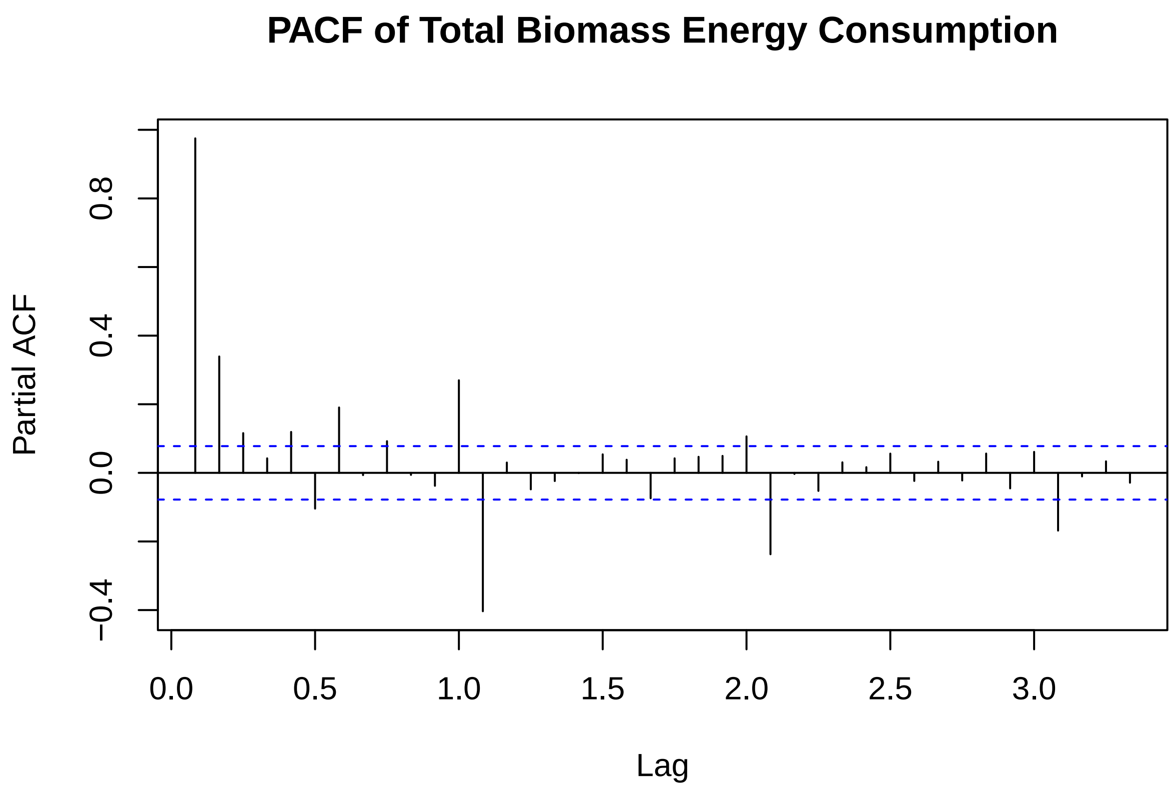
Their ACFs decrease as lags increase.

Biomass: Extremely persistent with very slow decay. Renewable: Very similar to the renewable, strong persistence. Hydroelectric: Moderate persistent and autocorrelation declines faster.

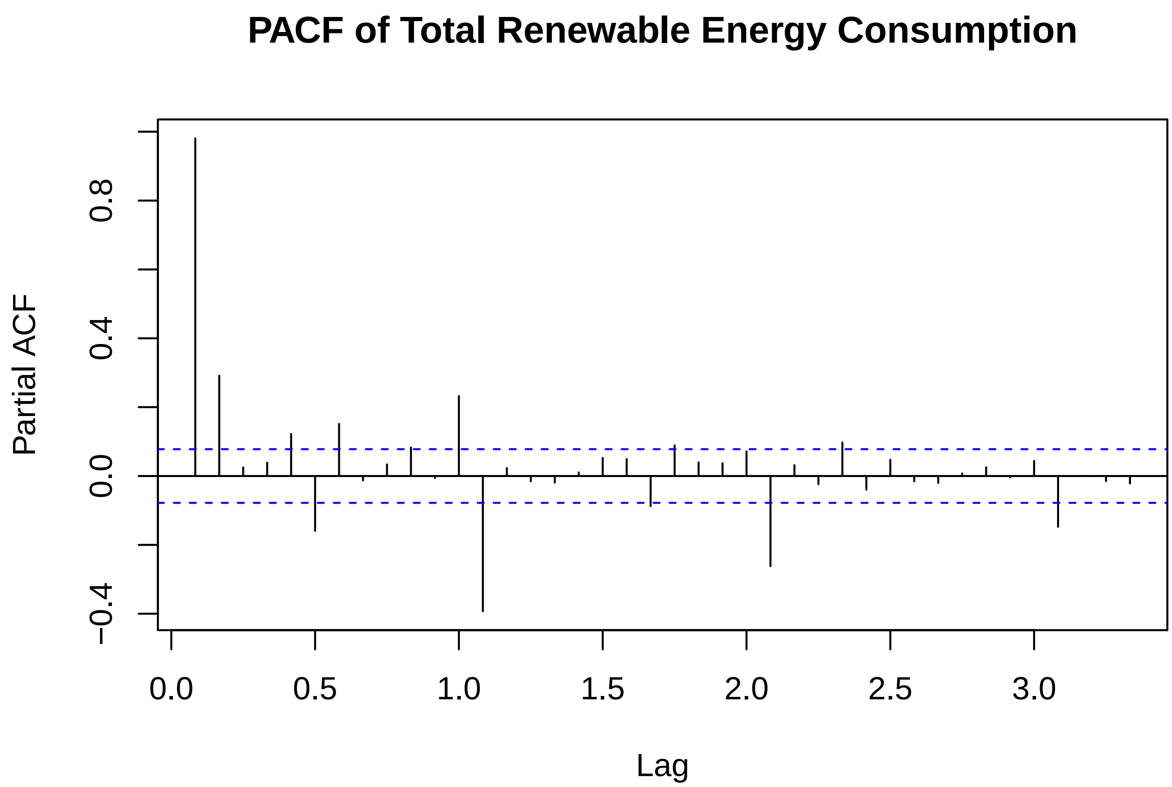
Question 7

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

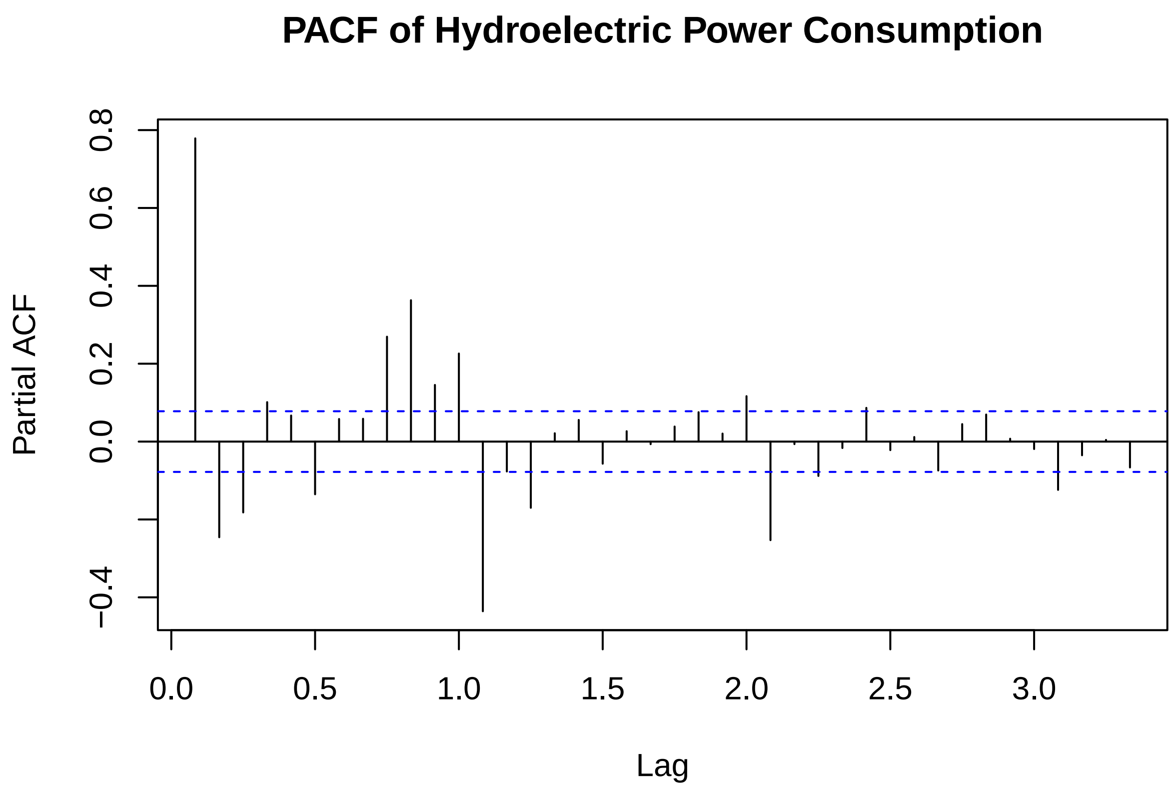
```
pacf_biomass <- pacf(ts_energy_data1[,1], lag.max = 40, plot = TRUE, main = ("PACF of Total Biomass Ene
```




```
pacf_renewable <- pacf(ts_energy_data1[,2], lag.max = 40, plot = TRUE, main = ("PACF of Total Renewable
```




```
pacf_hydroelectric <- pacf(ts_energy_data1[,3], lag.max = 40, plot = TRUE, main = ("PACF of Hydroelectri
```



In ACF, autocorrelations remain large for many lags, especially in Renewable and Biomass plots. In PACF, however, Lag 1 is very large and significant while most other lags are small and insignificant.