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

Assignment 2 - Due date 01/23/25

Rachael Stephan

## 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_Sp24.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); library(tseries); library(tidyverse)
library(readxl); library(openxlsx); library(lubridate)
library(knitr)
```

##Figure Theme

Create and set a cohesive theme for any figures.

```
#set theme
mytheme <- theme_bw(base_size = 10)+
  theme(axis.title = element_text(size = 10, hjust = 0.5),
        plot.title.position = "panel",
        panel.border = element_rect(colour = "black", fill = NA, linewidth = 0.25),
        plot.caption = element_text(hjust = 0),
        legend.box = "vertical",
        legend.location = "plot",
        axis.gridlines = element_line(color = "grey", linewidth = 0.25),
        axis.ticks = element_line(color = "black", linewidth = 0.5))
theme_set(mytheme)
```

## 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 2023 Monthly Energy Review. The spreadsheet is ready to be used. You will also find a *.csv* version of the data "Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source-Edit.csv". You may use the function *read.table*() to import the *.csv* data in R. Or refer to the file "M2_ImportingData_CSV_XLSX.Rmd" in our Lessons folder for functions that are better suited for importing the *.xlsx*.

```
#Importing data set
renewable_e_prod_consump <- read.xlsx("./Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_
                                      sheet = "Monthly Data",
                                      startRow = 13,
                                      colNames = FALSE)

#get column names
col_units <-
  read.xlsx("./Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source.xlsx",
            rows = 11:12,
            sheet="Monthly Data",
            colNames=FALSE)

#set col names
colnames(renewable_e_prod_consump) <- col_units[1,]

#fix dates
renewable_e_prod_consump$Month <- as_date(renewable_e_prod_consump$Month, origin = "1900-01-01")
renewable_e_prod_consump$Month <- paste(month(renewable_e_prod_consump$Month,
                                              label = TRUE,
                                              abbr = TRUE),
                                        year(renewable_e_prod_consump$Month))

#check data set
kable(head(renewable_e_prod_consump[,c(1:7)]),
      caption = "First few rows of the imported timeseries data")
kable(head(renewable_e_prod_consump[,c(1,8:14)]))
str(renewable_e_prod_consump)
```

Table 1: First few rows of the imported timeseries data

| Month | Wood Energy Production | Biofuels Production | Total Biomass Energy Production | Total Renewable Energy Production | Hydroelectric Power Consumption | Geothermal Energy Consumption |
|---|---|---|---|---|---|---|
| Jan 1973 | 129.630 | Not Available | 129.787 | 219.839 | 89.562 | 0.490 |
| Feb 1973 | 117.194 | Not Available | 117.338 | 197.330 | 79.544 | 0.448 |
| Mar 1973 | 129.763 | Not Available | 129.938 | 218.686 | 88.284 | 0.464 |
| Apr 1973 | 125.462 | Not Available | 125.636 | 209.330 | 83.152 | 0.542 |

| Month | Wood Energy Production | Biofuels Production | Total Biomass Energy Production | Total Renewable Energy Production | Hydroelectric Power Consumption | Geothermal Energy Consumption |
|---|---|---|---|---|---|---|
| May 1973 | 129.624 | Not Available | 129.834 | 215.982 | 85.643 | 0.505 |
| Jun 1973 | 125.435 | Not Available | 125.611 | 208.249 | 82.060 | 0.579 |

| Month | Solar Energy Consumption | Wind Energy Consumption | Wood Energy Consumption | Waste Energy Consumption | Biofuels Consumption | Total Biomass Energy Consumption | Total Renewable Energy Consumption |
|---|---|---|---|---|---|---|---|
| Jan 1973 | Not Available | Not Available | 129.630 | 0.157 | Not Available | 129.787 | 219.839 |
| Feb 1973 | Not Available | Not Available | 117.194 | 0.144 | Not Available | 117.338 | 197.330 |
| Mar 1973 | Not Available | Not Available | 129.763 | 0.176 | Not Available | 129.938 | 218.686 |
| Apr 1973 | Not Available | Not Available | 125.462 | 0.174 | Not Available | 125.636 | 209.330 |
| May 1973 | Not Available | Not Available | 129.624 | 0.210 | Not Available | 129.834 | 215.982 |
| Jun 1973 | Not Available | Not Available | 125.435 | 0.176 | Not Available | 125.611 | 208.249 |

```
'data.frame':   621 obs. of  14 variables:
 $ Month                            : chr  "Jan 1973" "Feb 1973" "Mar 1973" "Apr 1973" ...
 $ Wood Energy Production           : num  130 117 130 125 130 ...
 $ Biofuels Production              : chr  "Not Available" "Not Available" "Not Available" "Not Availa
 $ Total Biomass Energy Production  : num  130 117 130 126 130 ...
 $ Total Renewable Energy Production : num  220 197 219 209 216 ...
 $ Hydroelectric Power Consumption  : num  89.6 79.5 88.3 83.2 85.6 ...
 $ Geothermal Energy Consumption    : num  0.49 0.448 0.464 0.542 0.505 0.579 0.614 0.579 0.49 0.578 .
 $ Solar Energy Consumption         : chr  "Not Available" "Not Available" "Not Available" "Not Availa
 $ Wind Energy Consumption          : chr  "Not Available" "Not Available" "Not Available" "Not Availa
 $ Wood Energy Consumption          : num  130 117 130 125 130 ...
 $ Waste Energy Consumption         : num  0.157 0.144 0.176 0.174 0.21 0.176 0.17 0.184 0.178 0.2 ...
 $ Biofuels Consumption             : chr  "Not Available" "Not Available" "Not Available" "Not Availa
 $ Total Biomass Energy Consumption  : num  130 117 130 126 130 ...
 $ Total Renewable Energy Consumption: num  220 197 219 209 216 ...
```

## 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.*

```
#select for columns of interest
energy_dataset <- renewable_e_prod_consump %>%
  select(`Total Biomass Energy Production`,
         `Total Renewable Energy Production`,
         `Hydroelectric Power Consumption`)

#get first few rows of each column
kable(head(energy_dataset),
      caption = "First few rows of the selected timeseries for analysis")
```

Table 3: First few rows of the selected timeseries for analysis

| Total Biomass Energy Production | Total Renewable Energy Production | Hydroelectric Power Consumption |
|---:|---:|---:|
| 129.787 | 219.839 | 89.562 |
| 117.338 | 197.330 | 79.544 |
| 129.938 | 218.686 | 88.284 |
| 125.636 | 209.330 | 83.152 |
| 129.834 | 215.982 | 85.643 |
| 125.611 | 208.249 | 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().*

```
#create a time series object specifying the start and frequency of ts object
energy_dataset_ts <- ts(energy_dataset,
                        start=c(1973,1),
                        frequency=12)
```

## Question 3

*Compute mean and standard deviation for these three series.*

```
#calculate the mean for each ts
energy_dataset_mean <- sapply(energy_dataset_ts, mean)

#calculate the std. dev for each ts
energy_dataset_std <- sapply(energy_dataset_ts, sd)

#display the mean and standard deviation
kable(cbind("Mean (Trillion Btu)" = energy_dataset_mean,
            "Standard Deviation (Trillion Btu)" = energy_dataset_std),
      caption = "The mean and standard deviation of each selected time series")
```

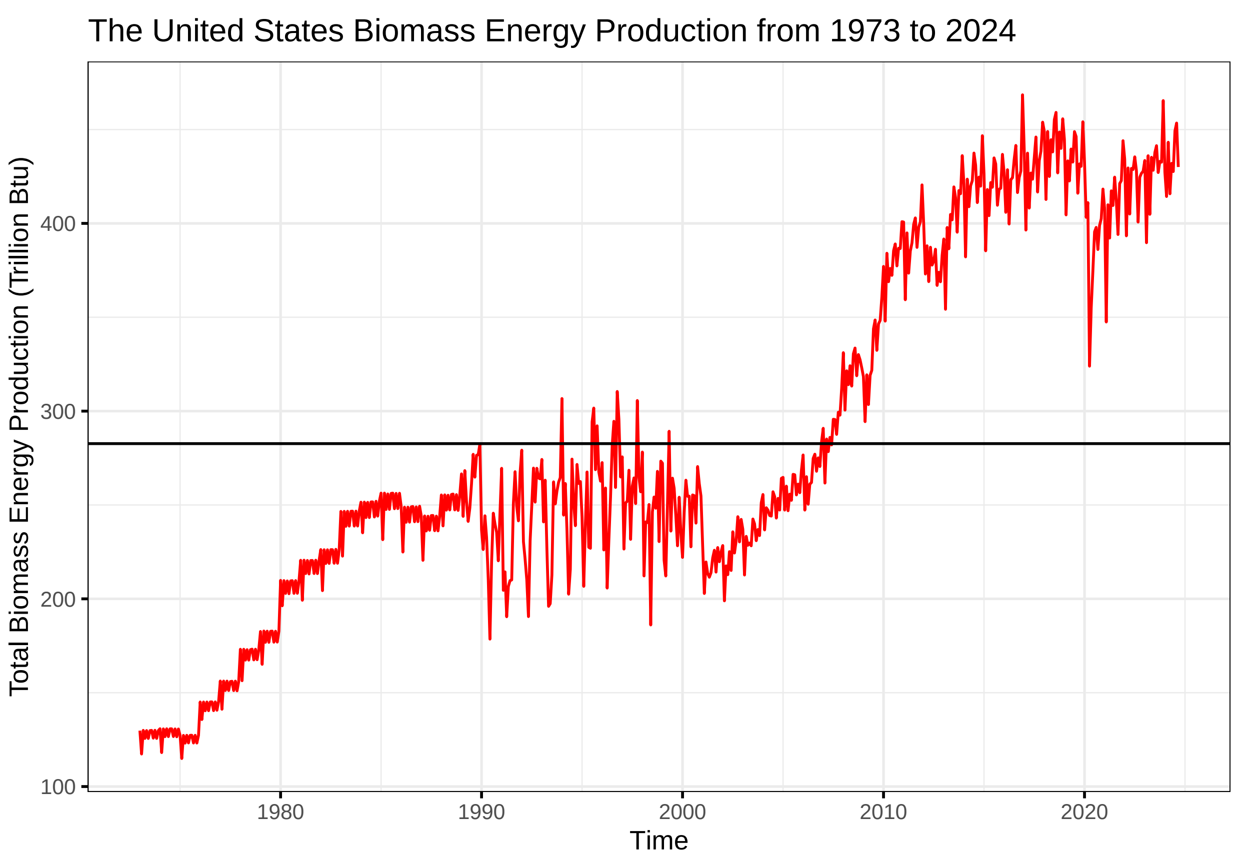
Table 4: The mean and standard deviation of each selected time series

| | Mean (Trillion Btu) | Standard Deviation (Trillion Btu) |
|---|---|---|
| Total Biomass Energy Production | 282.67785 | 94.05815 |
| Total Renewable Energy Production | 402.01667 | 143.79270 |
| Hydroelectric Power Consumption | 79.55371 | 14.10737 |

## Question 4

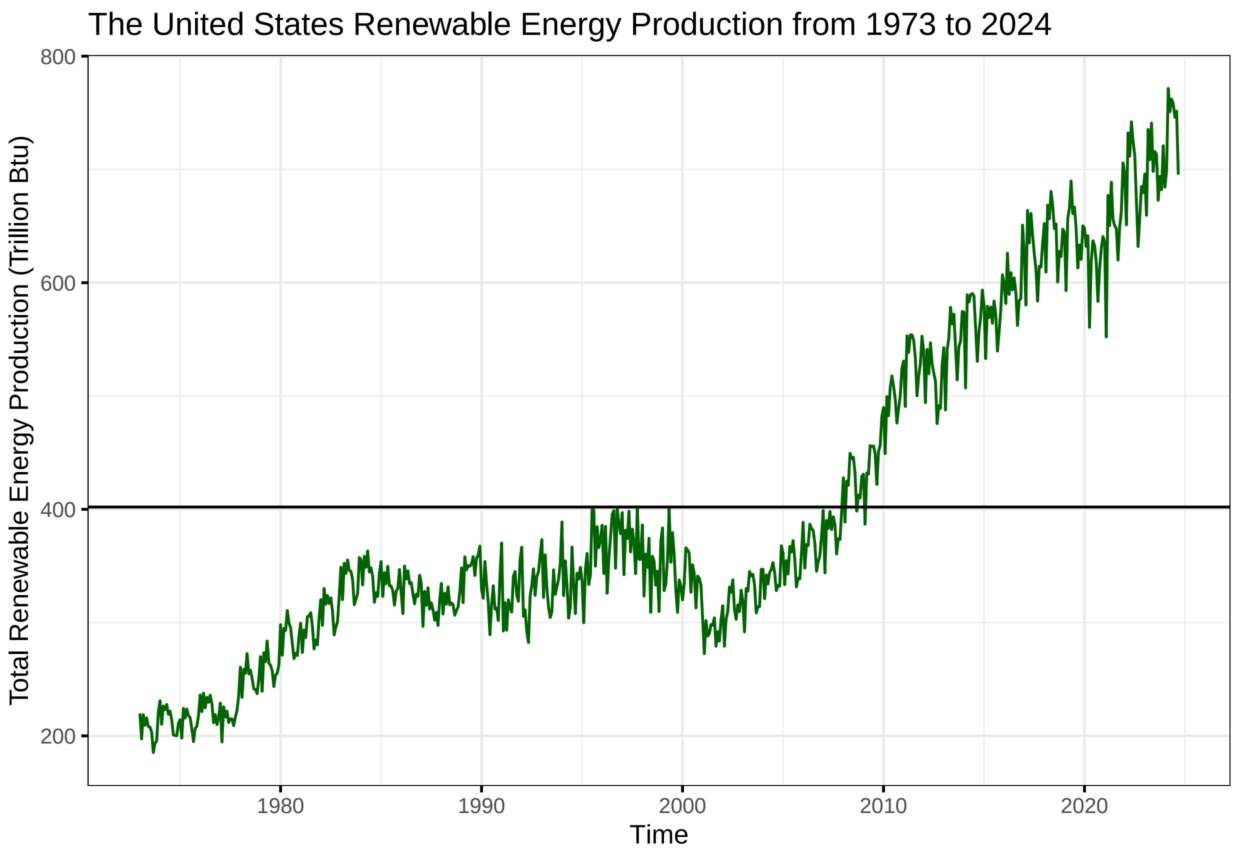
*Display and interpret the time series plot for each of these variables. Try to make your plot as informative as possible by writing titles, labels, etc. For each plot add a horizontal line at the mean of each series in a different color.*

```
#plot biomass ts
autoplot(energy_dataset_ts[,1], col = "red")+
  geom_abline(slope = 0, intercept = energy_dataset_mean[1])+
  labs(y = paste(col_units[1,4],col_units[2,4], sep = " "),
       title = "The United States Biomass Energy Production from 1973 to 2024")
```

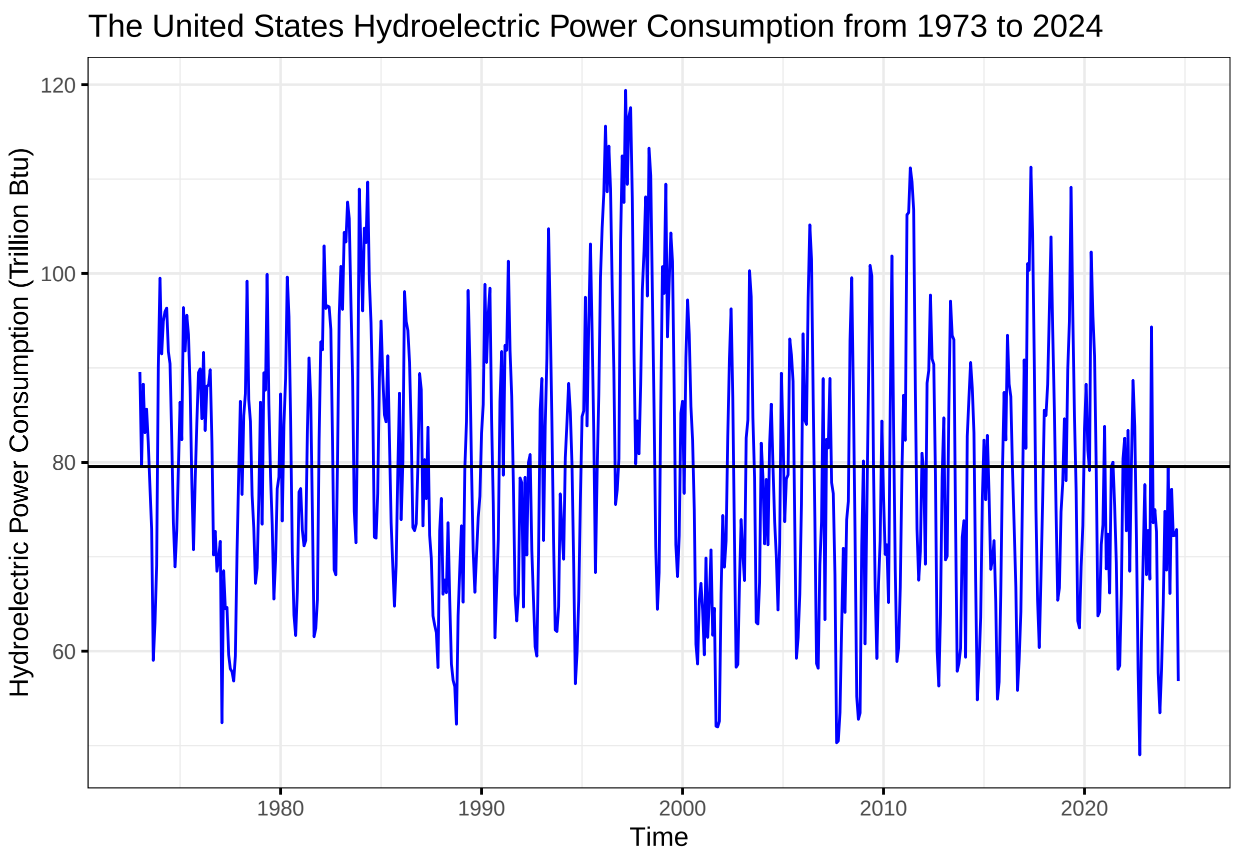



```
#plot renewable ts
autoplot(energy_dataset_ts[,2], col ="darkgreen")+
  geom_abline(slope = 0, intercept = energy_dataset_mean[2])+
```

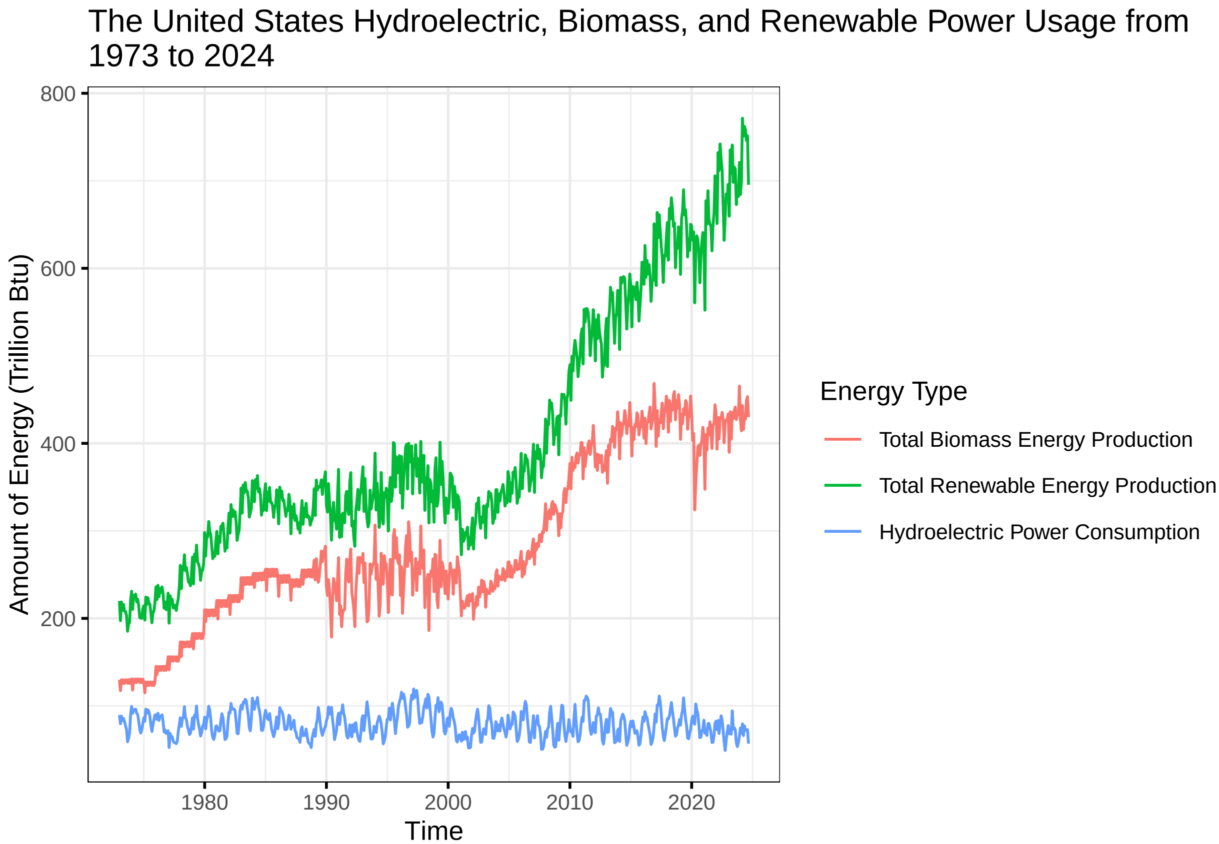
```r
  labs(y = paste(col_units[1,5],col_units[2,5], sep = " "),
       title = "The United States Renewable Energy Production from 1973 to 2024")
```




```r
#plot renewable ts
autoplot(energy_dataset_ts[,3], col ="blue")+
  geom_abline(slope = 0, intercept = energy_dataset_mean[3])+
  labs(y = paste(col_units[1,6],col_units[2,6], sep = " "),
       title = "The United States Hydroelectric Power Consumption from 1973 to 2024")
```




```
#plot all ts
autoplot(energy_dataset_ts)+
  labs(y = paste("Amount of Energy",col_units[2,6], sep = " "),
       title = "The United States Hydroelectric, Biomass, and Renewable Power Usage from\n1973 to 2024"
       colour = "Energy Type")
```




Both the total biomass energy production and the total renewable energy production increases over time. These two time series show similar trends to each other in which they go through a period of linear increase, followed by a stable period, and then another upward linear trend. The biomass timeseries appears to have a steep reduction in energy production in 2020 but then continues with the positive trend afterwards. The renewable energy timeseries has a slight decrease in 2020 as well, but it is not as pronounced. Th

There appears to be periods in the biomass timeseries where the magnitude of the seasonality changes, with greater magnitudes occurring between approximately 1990 to 2000 and from 2015 onwards. These time periods are also the periods where the trend of biomass production is more stable. Conversely, the magnitude of the seasonality in the renewable energy production timeseries appears to remain stable across the timeseries.

The hydroelectric power consumption appears to remain stable over time. The peaks and valleys (i.e., seasonality) of the time series also appear to be relatively similar in magnitude over time.

## Question 5

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

```
energy_dataset_corr <- cor(energy_dataset_ts)

kable(energy_dataset_corr,
      caption = "The correlation coefficients for the biomass, renewable, and hydroelectric energy use '
```

Table 5: The correlation coefficients for the biomass, renewable, and hydroelectric energy use timeseries in the United States from 1973 to 2024

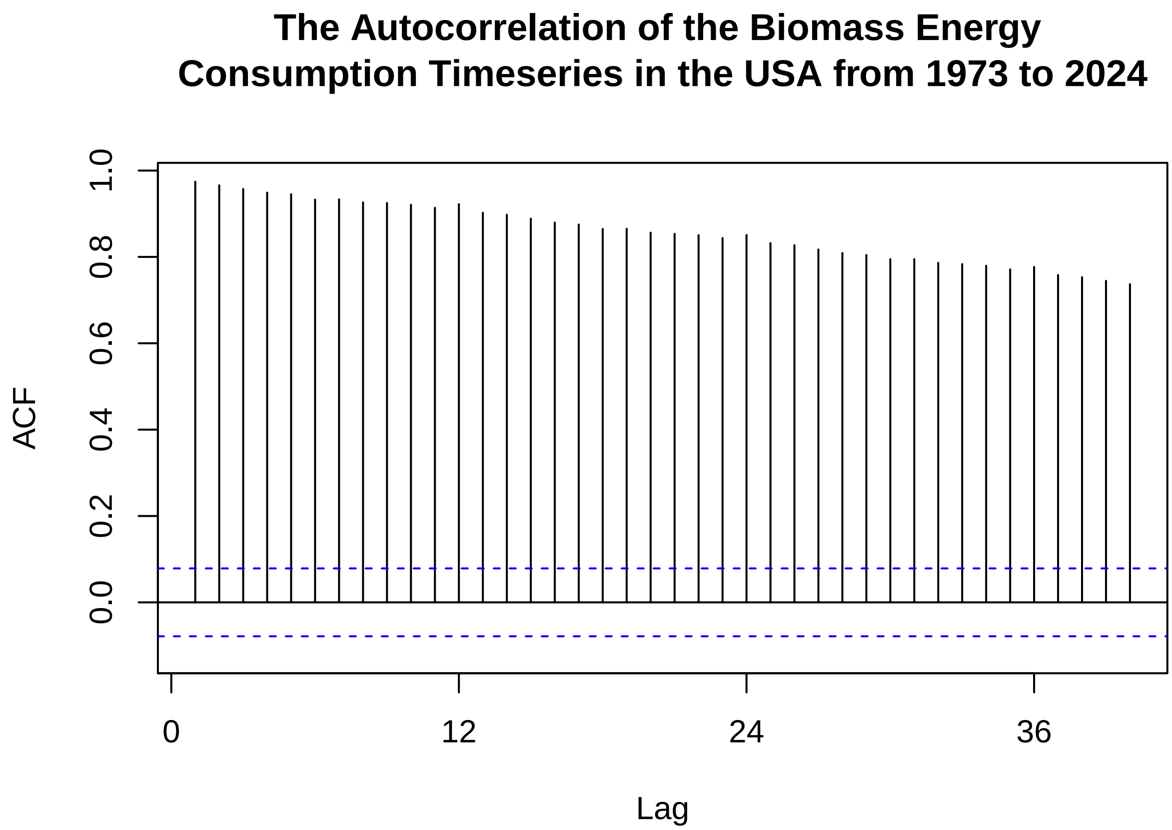
| | Total Biomass Energy Production | Total Renewable Energy Production | Hydroelectric Power Consumption |
|---|---|---|---|
| Total Biomass Energy Production | 1.0000000 | 0.9678137 | -0.1142927 |
| Total Renewable Energy Production | 0.9678137 | 1.0000000 | -0.0291610 |
| Hydroelectric Power Consumption | -0.1142927 | -0.0291610 | 1.0000000 |

The total biomass energy production and the total renewable energy production are closely correlated, as indicated by a high absolute value for the correlation coefficient. None of the other timeseries are closely correlated with each other, as shown by low absolute values for their correlation coefficients. This results aligns with the observations made previously.

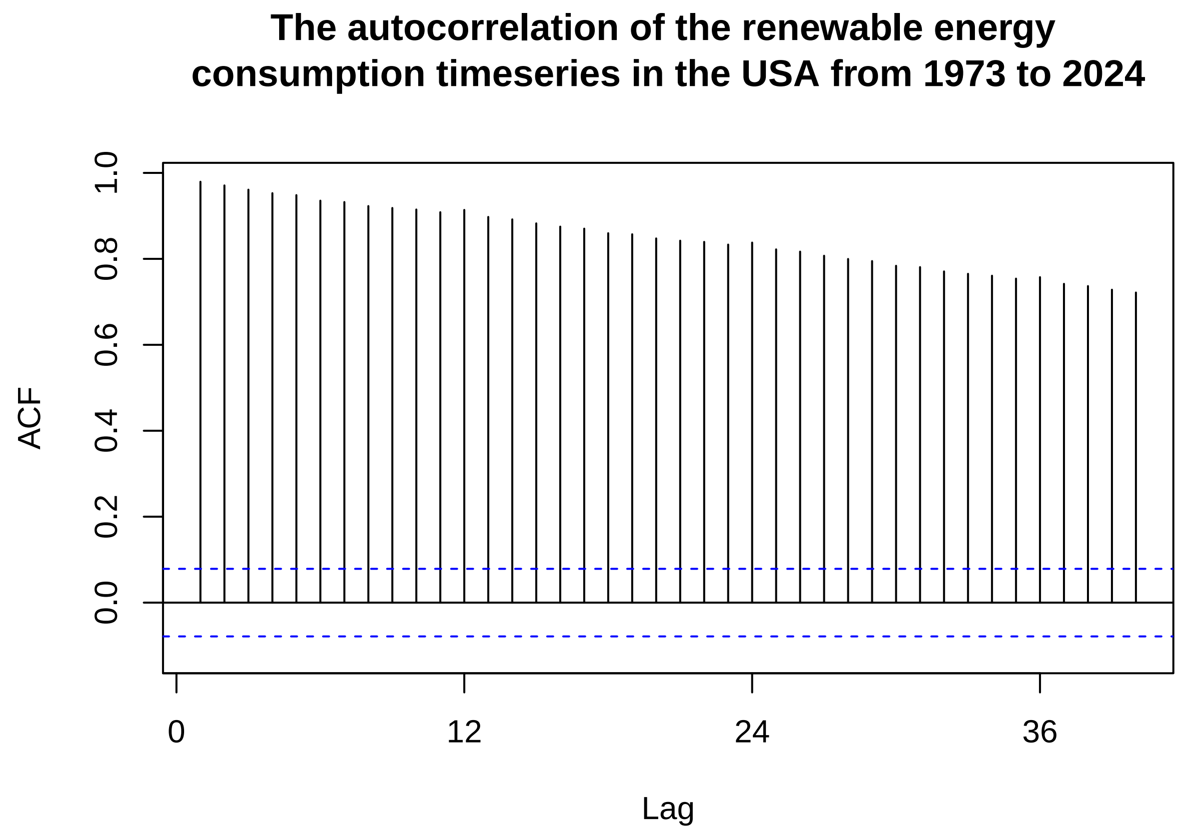
## Question 6

*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?*

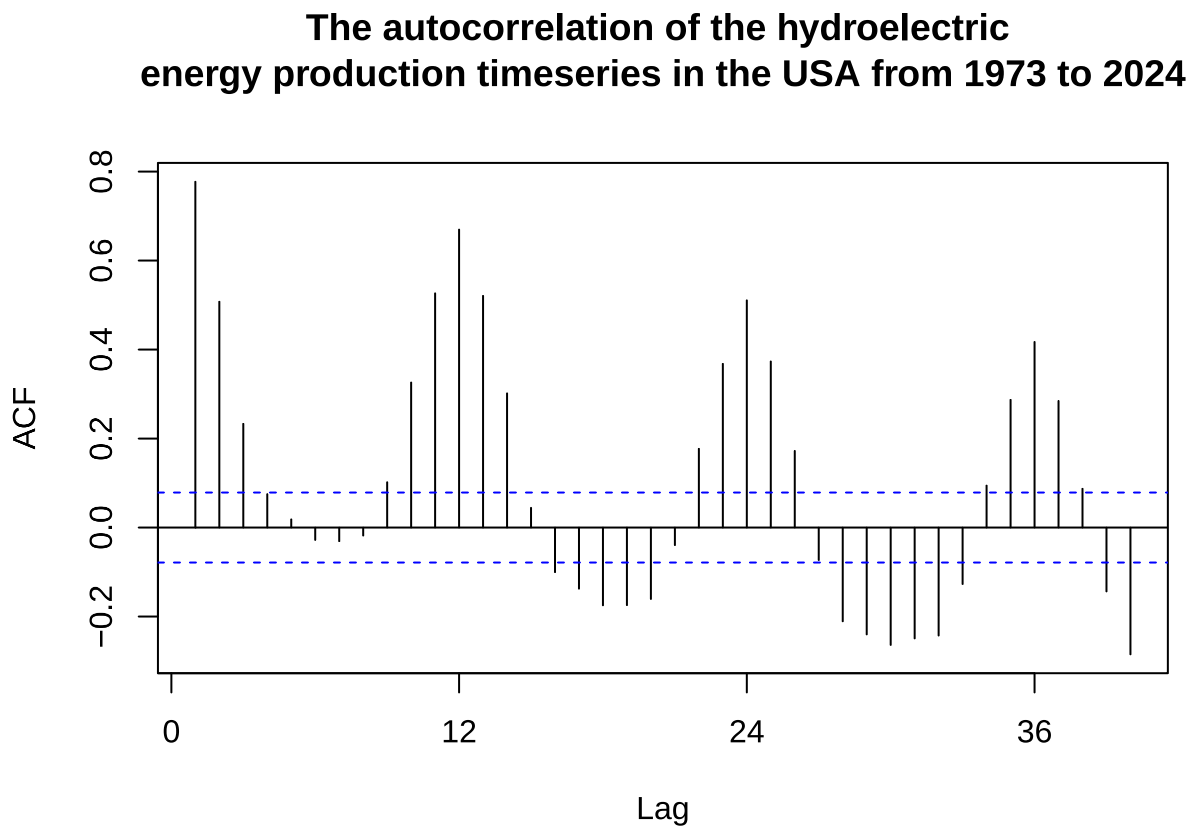
```
energy_dataset_acf_biomass <- Acf(energy_dataset_ts[,1],
                                  lag.max=40,
                                  type="correlation",
                                  plot=TRUE,
                                  main = "The Autocorrelation of the Biomass Energy
Consumption Timeseries in the USA from 1973 to 2024")
```




```
energy_dataset_acf_renewable <- Acf(energy_dataset_ts[,2],
                                    lag.max=40,
                                    type="correlation",
                                    plot=TRUE,
                                    main = "The autocorrelation of the renewable energy
consumption timeseries in the USA from 1973 to 2024")
```




```r
energy_dataset_acf_hydroelectric <- Acf(energy_dataset_ts[,3],
                                        lag.max=40,
                                        type="correlation",
                                        plot=TRUE,
                                        main = "The autocorrelation of the hydroelectric
energy production timeseries in the USA from 1973 to 2024")
```

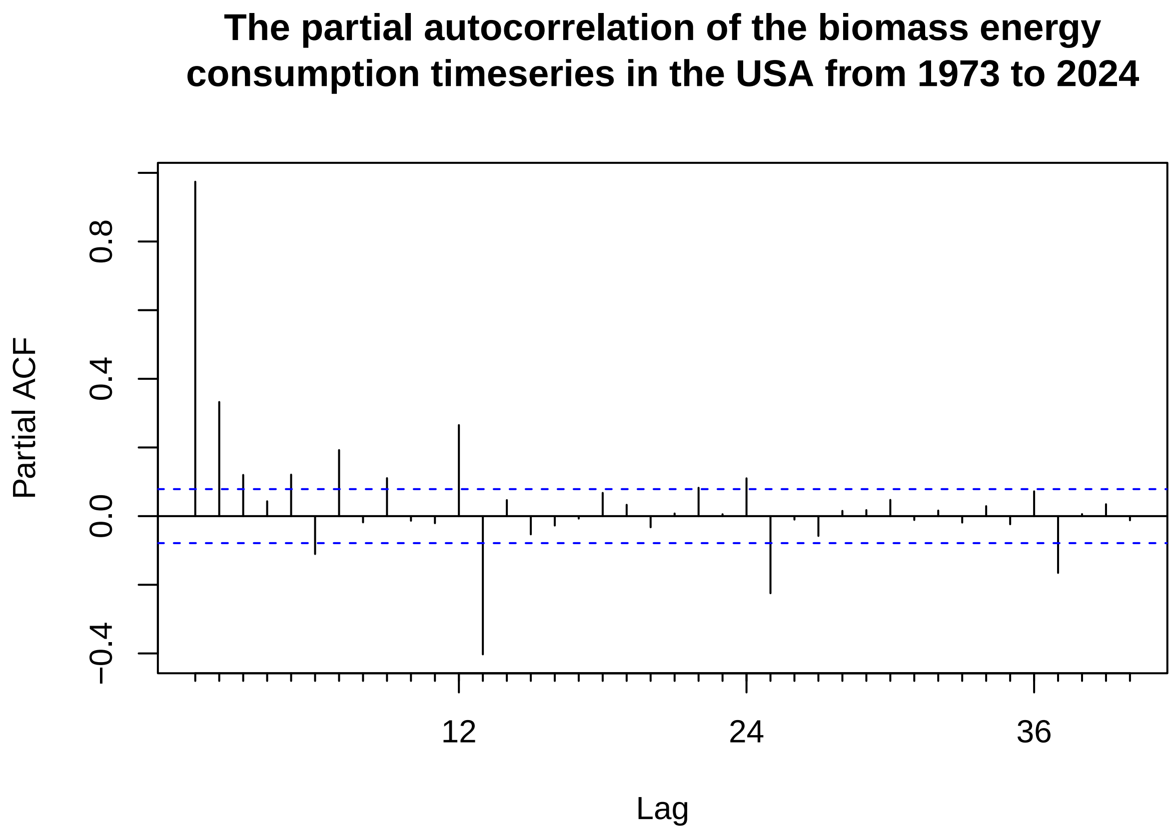



The renewable and biomass energy consumption autocorrelation plots show similar behaviour. Both time series start out with a highly positive autocorrelation and decreases at a relatively consistent rate with each lag. Both of these timeseries still maintain a high magnitude of correlation at 40 lags. The autocorrelation plot for the hydroelectric energy production shows different behaviour. This plot oscillates between higher and lower magnitude correlations, alternating between positive and negative correlations. This could indicate the presence of seasonality. As the lags progress, the maximum magnitudes of the positive autocorrelation decreases but the maximum magnitude of the negative autocorrelations increases.

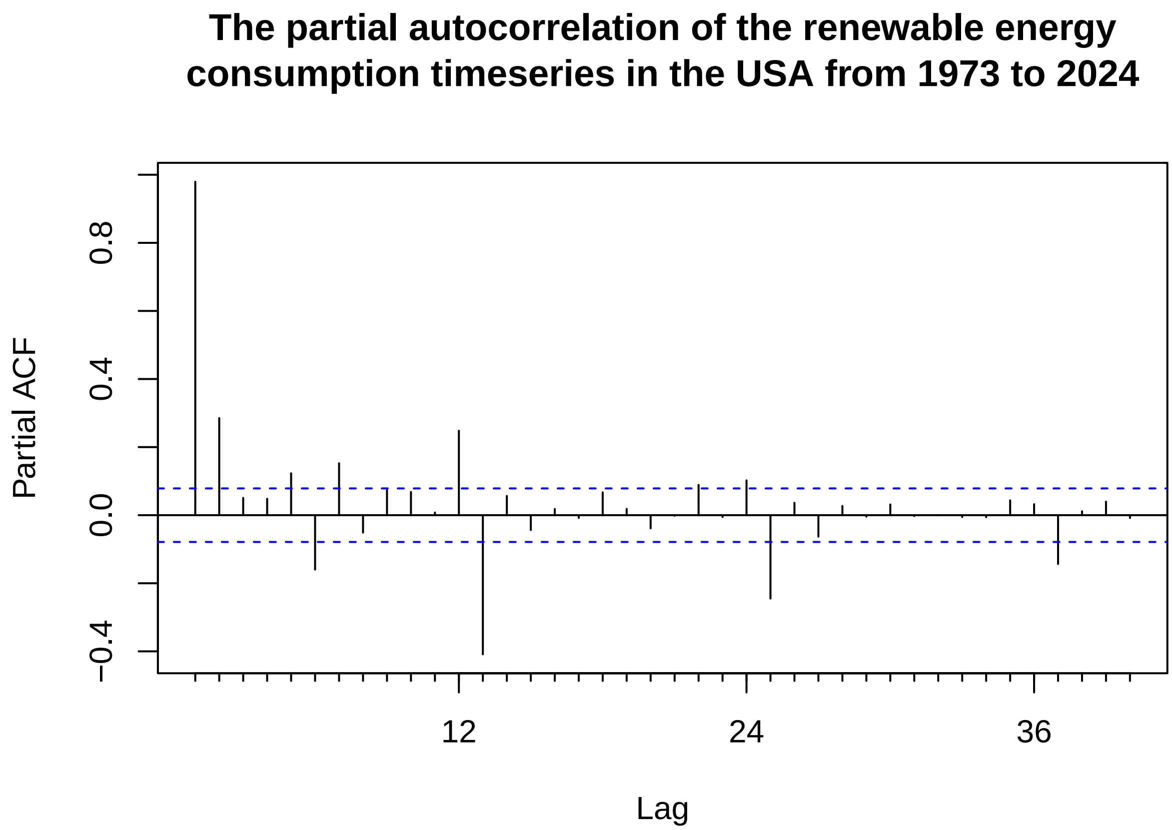
## Question 7

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

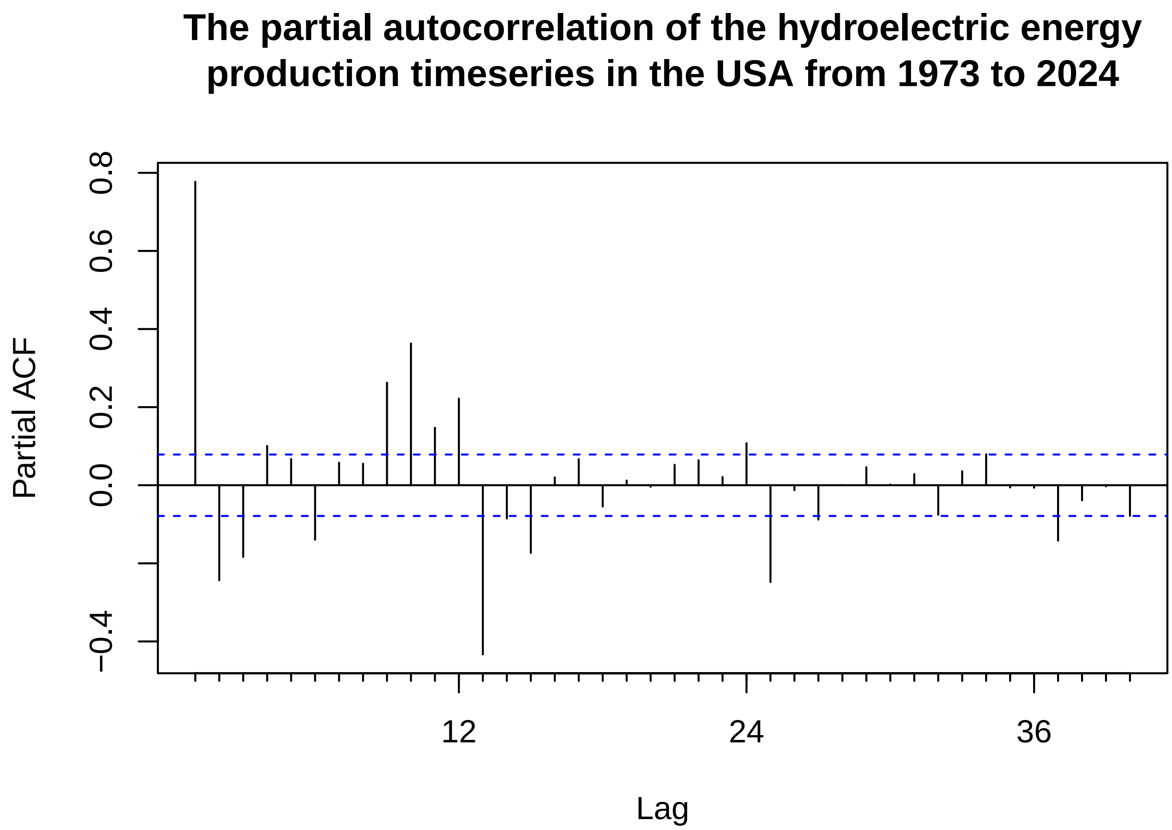
```
energy_dataset_pacf_biomass <- Pacf(energy_dataset_ts[,1],
                                    lag.max=40,
                                    plot=TRUE,
                                    main = "The partial autocorrelation of the biomass energy\nconsumptio
```




```
energy_dataset_pacf_renewable <- Pacf(energy_dataset_ts[,2],
                                      lag.max=40,
                                      plot=TRUE,
                                      main = "The partial autocorrelation of the renewable energy\nconsumpti
```




```
energy_dataset_pacf_hydroelectric <- Pacf(energy_dataset_ts[,3],
                                  lag.max=40,
                                  plot=TRUE,
                                  main = "The partial autocorrelation of the hydroelectric energy\nprodu
```




All of the partial autocorrelation plots show similar trends. There is a general decay in amplitude of all of the PACF plots. Only the highest amplitudes surpass the blue dotted line that indicates significance. The amount of times that exceed the blue line decreases with increasing lags. There is a smaller oscillating motion between positive and negative PACF that may coincide with seasonality in the data. The renewable energy and biomass energy PACF plots have similar magnitudes, and the hydroelectric energy PACF magnitudes were slightly less in comparison.