ENV 790.30 - Time Series Analysis for Energy Data | Spring 2024

Assignment 2 - Due date 02/25/24

Student Name

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)
```

```
## 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
```

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
library(readr)
dataset_raw <- read_csv("./Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source.csv", 
```

```
## Rows: 609 Columns: 14
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (5): Month, Biofuels Production, Solar Energy Consumption, Wind Energy C...
## dbl (9): Wood Energy Production, Total Biomass Energy Production, Total Rene...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
head(dataset_raw)
```

```
## # A tibble: 6 x 14
##   Month      Wood Energy Producti~1 `Biofuels Production` Total Biomass Energy~2
##   <chr>                       <dbl> <chr>                                  <dbl>
## 1 1973 Janu~                   130. Not Available                           130.
## 2 1973 Febr~                   117. Not Available                           117.
## 3 1973 March                   130. Not Available                           130.
## 4 1973 April                   125. Not Available                           126.
## 5 1973 May                     130. Not Available                           130.
## 6 1973 June                    125. Not Available                           126.
## # i abbreviated names: 1: `Wood Energy Production`,
## #   2: `Total Biomass Energy Production`
## # i 10 more variables: `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>, ...
```

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.

```
#take three columns out and createae  dataframe
dataset_final <- dataset_raw[,c(1,4:6)]
colnames(dataset_final)<-c("Date", "Biomass", "Renewables", "Hydro")
head(dataset_final)
```

```
## # A tibble: 6 x 4
##   Date          Biomass Renewables Hydro
##   <chr>           <dbl>      <dbl> <dbl>
## 1 1973 January     130.       220.  89.6
## 2 1973 February    117.       197.  79.5
## 3 1973 March       130.       219.  88.3
## 4 1973 April       126.       209.  83.2
## 5 1973 May         130.       216.  85.6
## 6 1973 June        126.       208.  82.1
```

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().

```
#Step 1: Converting the date column to ymd
library(lubridate)
```

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

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

```
DateColumn<-paste(dataset_final$Date, "01", sep = "-")
#converting column 1 to a yymmdd format by adding 01 as the "d"
DateColumn_ymd <- ymd(DateColumn)
#converting the above output to the ymd format using lubridate's ymd function
dataset_final$Date <- DateColumn_ymd
#switching the month column in the dataset to the ymd format made above
class(dataset_final$Date)
```

```
## [1] "Date"
```

```
#checking the class of the new Month column in the dataset
head(dataset_final) #checking the dataset
```

```
## # A tibble: 6 x 4
##   Date       Biomass Renewables Hydro
##   <date>       <dbl>      <dbl> <dbl>
## 1 1973-01-01    130.       220.  89.6
## 2 1973-02-01    117.       197.  79.5
## 3 1973-03-01    130.       219.  88.3
## 4 1973-04-01    126.       209.  83.2
## 5 1973-05-01    130.       216.  85.6
## 6 1973-06-01    126.       208.  82.1
```

```
#Step 2: Converting the dataset into a time series object
dataset_final_ts <- ts(dataset_final, start = 1973, frequency = 12)
head(dataset_final_ts)
```

```
##           Date Biomass Renewables  Hydro
## Jan 1973 1096 129.787    219.839 89.562
## Feb 1973 1127 117.338    197.330 79.544
## Mar 1973 1155 129.938    218.686 88.284
## Apr 1973 1186 125.636    209.330 83.152
## May 1973 1216 129.834    215.982 85.643
## Jun 1973 1247 125.611    208.249 82.060
```

Question 3

Compute mean and standard deviation for these three series.

```
Mean_Biomass <- mean(dataset_final_ts[,2])
Mean_Biomass
```

```
## [1] 279.8046
```

```
SD_Biomass <- sd(dataset_final_ts[,2])
SD_Biomass
```

```
## [1] 92.66504
```

```
Mean_Renewables <- mean(dataset_final_ts[,3])
Mean_Renewables
```

```
## [1] 395.7213
```

```
SD_Renewables <- sd(dataset_final_ts[,3])
SD_Renewables
```

```
## [1] 137.7952
```

```
Mean_Hydro <- mean(dataset_final_ts[,4])
Mean_Hydro
```

```
## [1] 79.73071
```

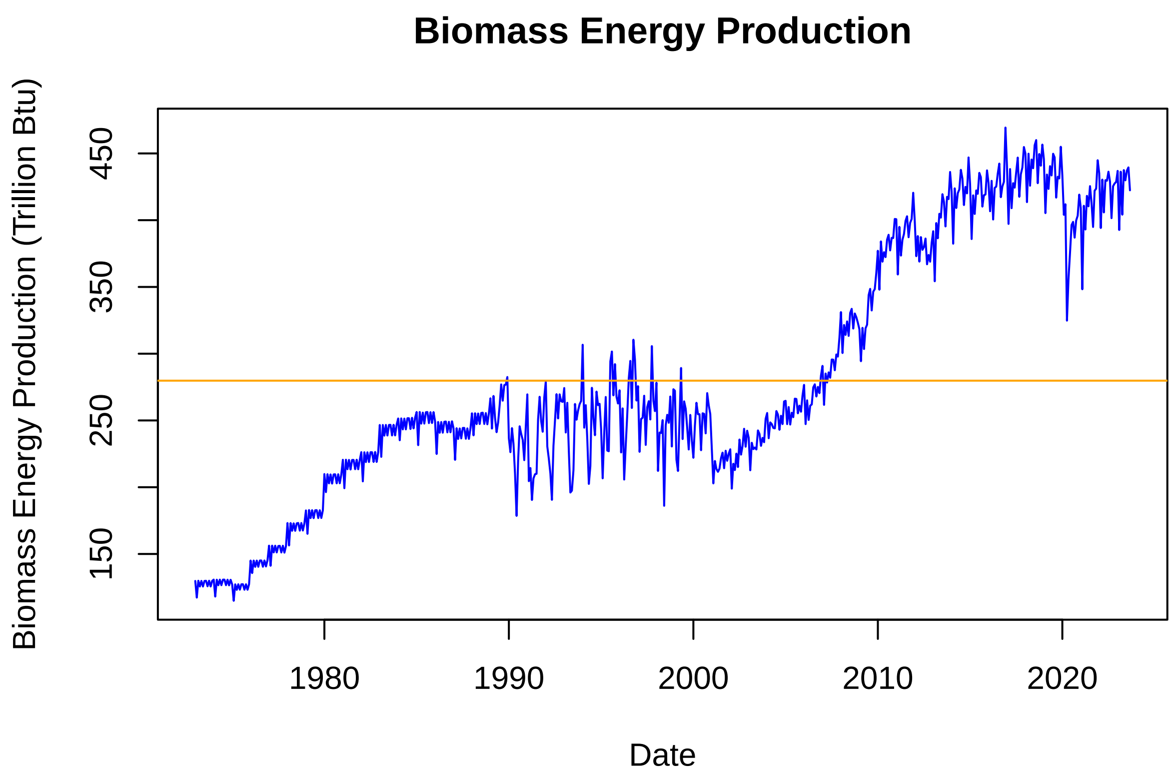
```
SD_Hydro <- sd(dataset_final_ts[,4])
SD_Hydro
```

```
## [1] 14.14734
```

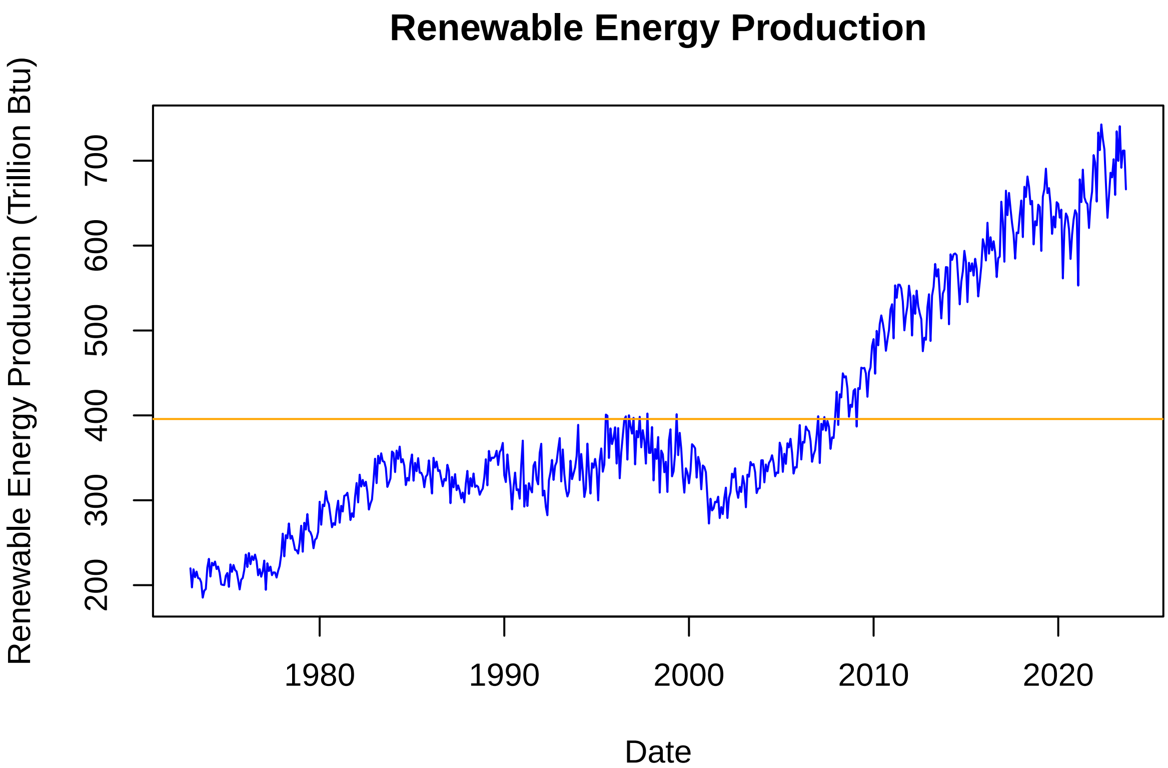
Question 4

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.

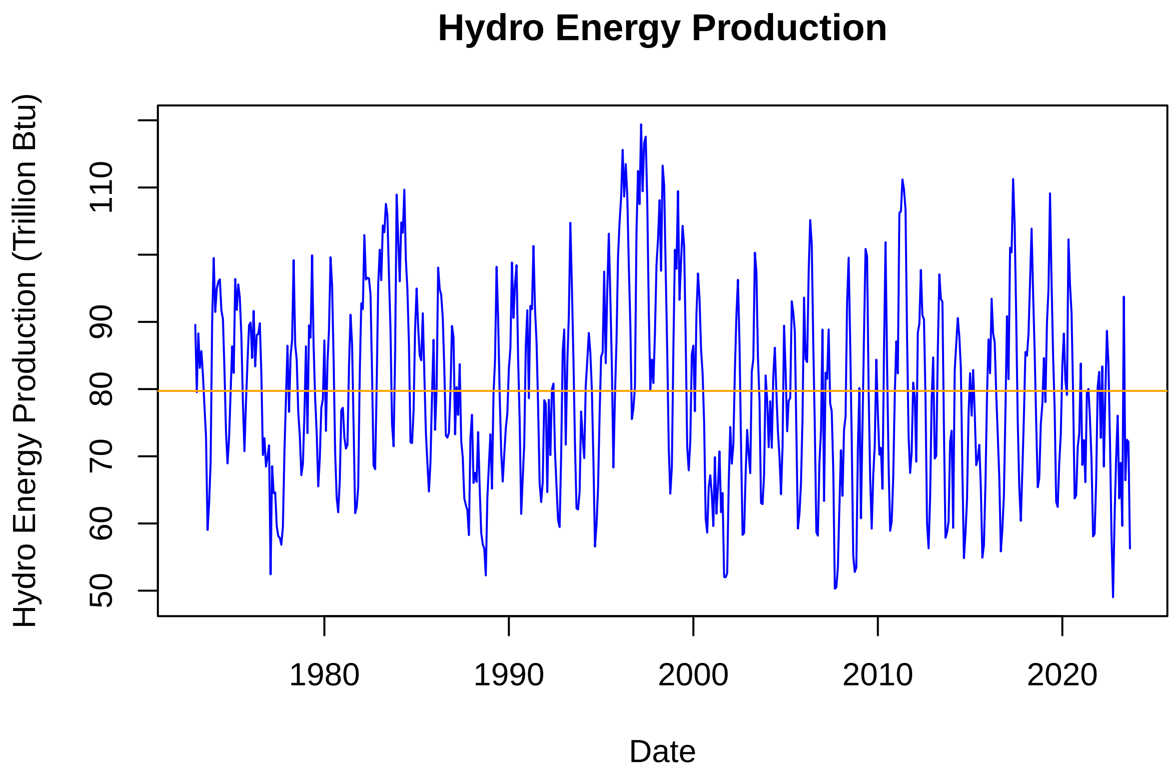
```r
#Plot for Biomass
Plot_Biomass <- plot(dataset_final_ts[,"Biomass"],
                main = "Biomass Energy Production",
                xlab = "Date", ylab = "Biomass Energy Production (Trillion Btu)",
                col="Blue")+
  abline(h=mean(Mean_Biomass), col = "Orange")
```




```r
#Plot for Renewables
Plot_Renewables <- plot(dataset_final_ts[,"Renewables"],
                main = "Renewable Energy Production",
                xlab = "Date", ylab = "Renewable Energy Production (Trillion Btu)",
                col = "Blue")+
  abline(h=mean(Mean_Renewables), col = "Orange")
```




```r
#Plot for Hydro
Plot_Hydro <- plot(dataset_final_ts[,"Hydro"],
                   main = "Hydro Energy Production",
                   xlab = "Date", ylab = "Hydro Energy Production (Trillion Btu)",
                   col = "Blue")+
  abline(h=mean(Mean_Hydro), col = "Orange")
```



Question 5

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

```
Cor_Biomass_RE <- cor(dataset_final_ts[,"Biomass"], dataset_final_ts[,"Renewables"])
Cor_Biomass_RE
```

```
## [1] 0.9707462
```

```
Cor_Biomass_Hydro <- cor(dataset_final_ts[,"Biomass"], dataset_final_ts[,"Hydro"])
Cor_Biomass_Hydro
```

```
## [1] -0.09656318
```

```
Cor_Hydro_RE <- cor(dataset_final_ts[,"Hydro"], dataset_final_ts[,"Renewables"])
Cor_Hydro_RE
```

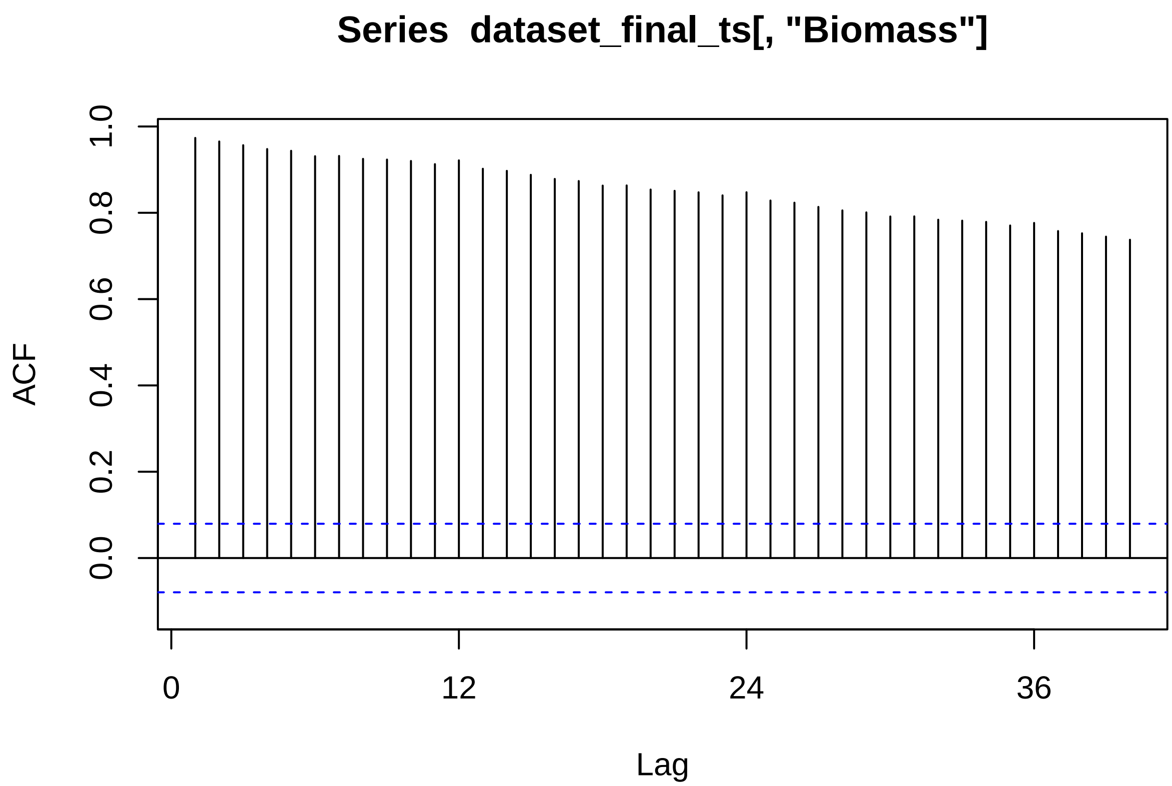
```
## [1] -0.001768629
```

```
#biomass and RE are very strongly correlated at 0.97, but hydro is almost
#uncorrelated with biomass and RE at -0.09 and -0.0017 respectively.
```

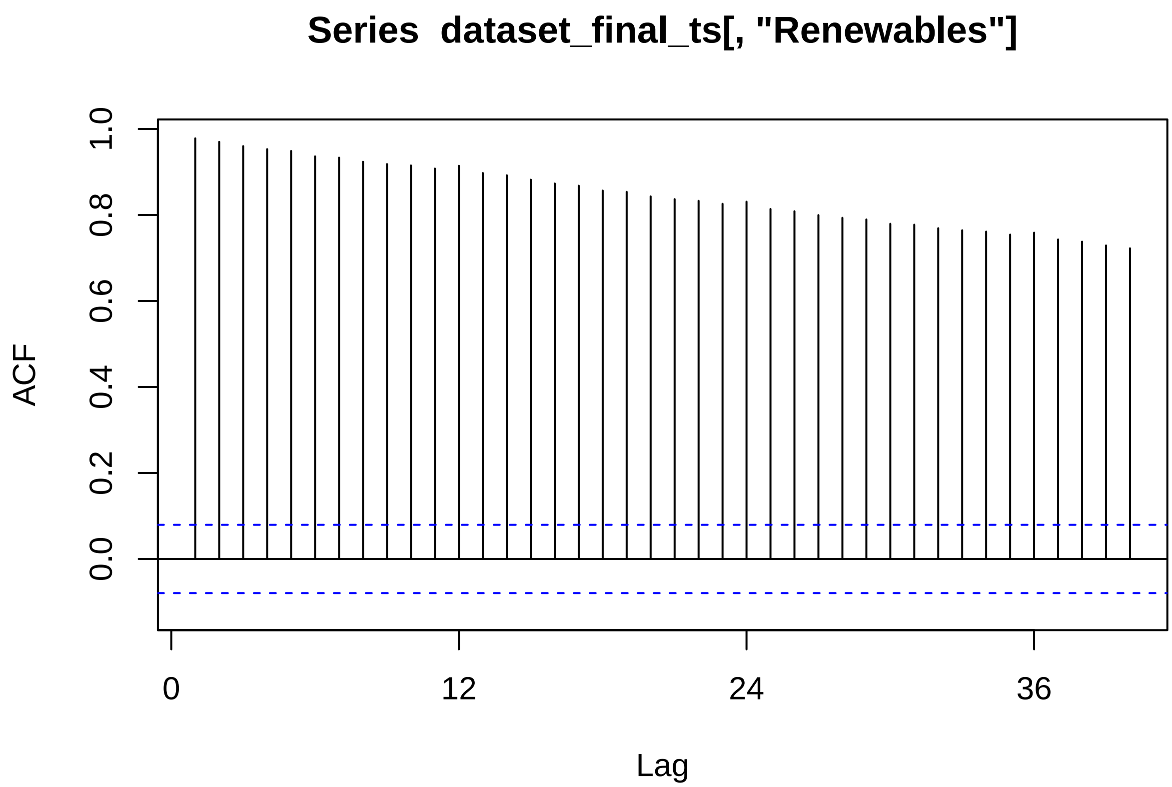
Question 6

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?

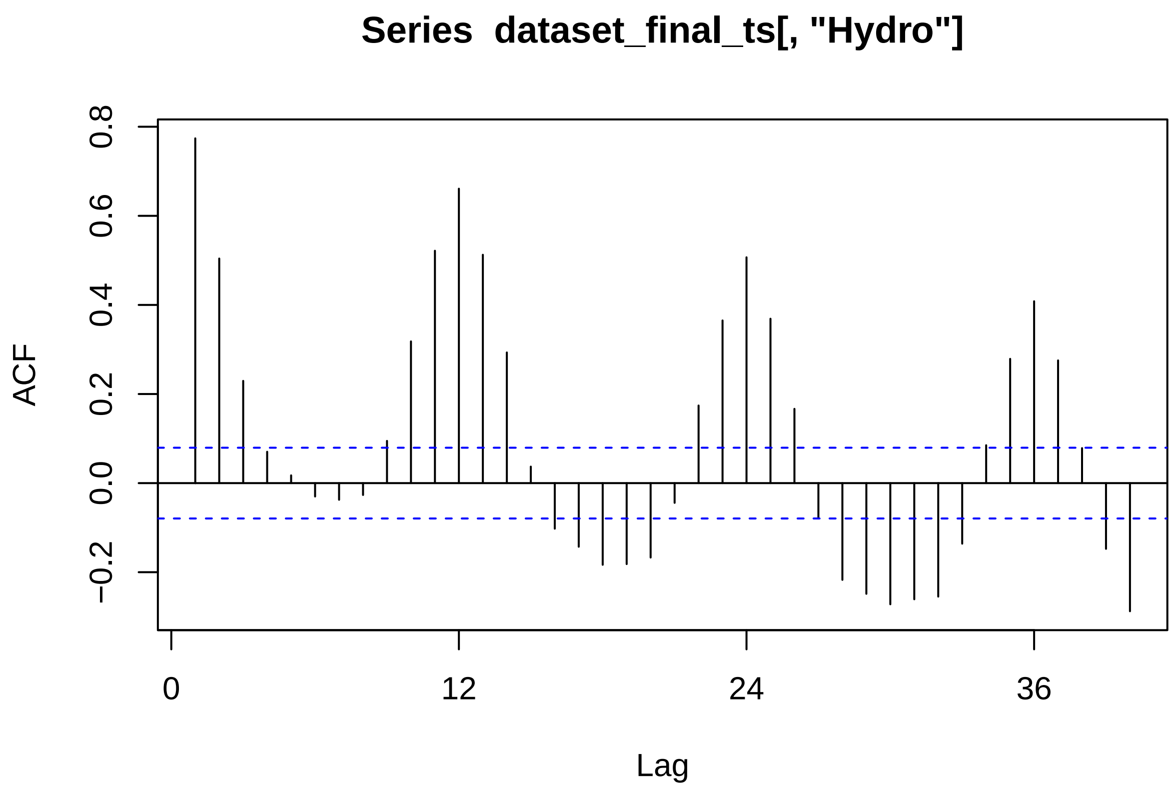
```r
library(forecast)
Acf(dataset_final_ts[,"Biomass"], lag.max = 40)
```




```r
Acf(dataset_final_ts[,"Renewables"], lag.max = 40)
```




```r
Acf(dataset_final_ts[,"Hydro"], lag.max = 40)
```

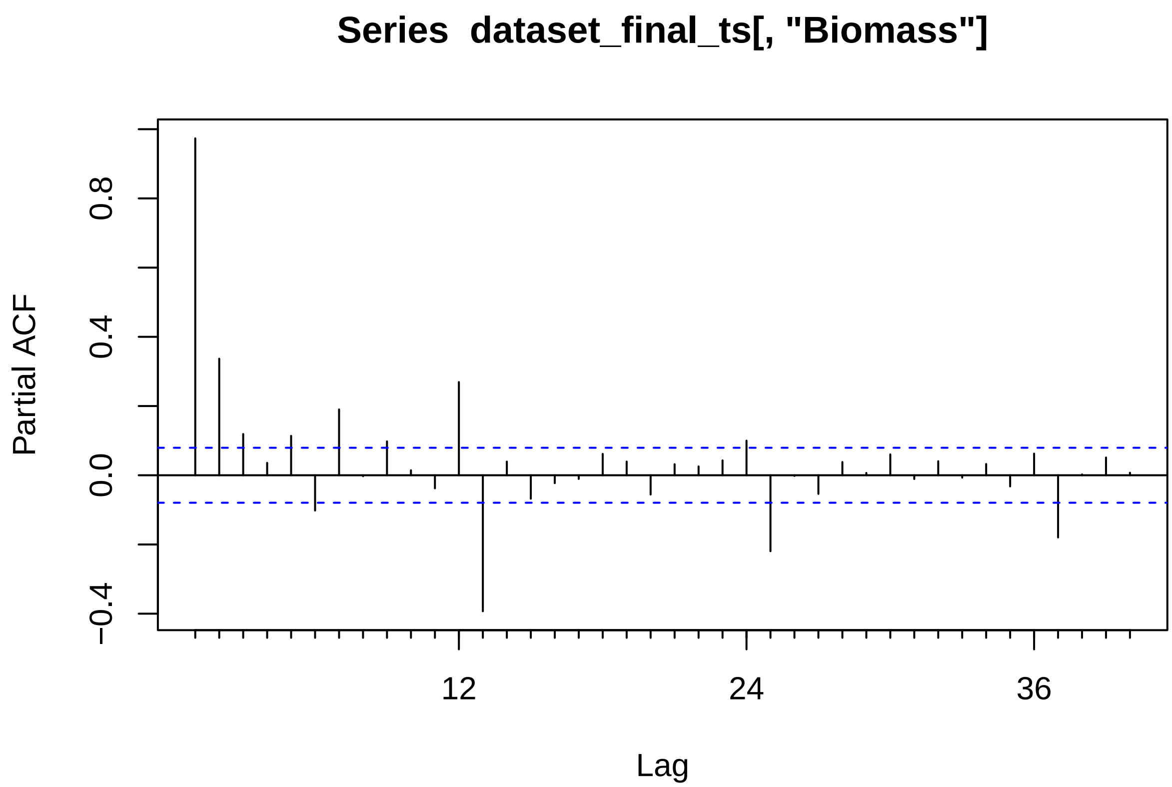



```
#Biomass and Renewables have similar behavior which is that the autocorrelation
#decreases  gradually over time/ lags. However, hydro follows a completely
#different pattern wherein it's autocorrelation follows a cyclical pattern.
```

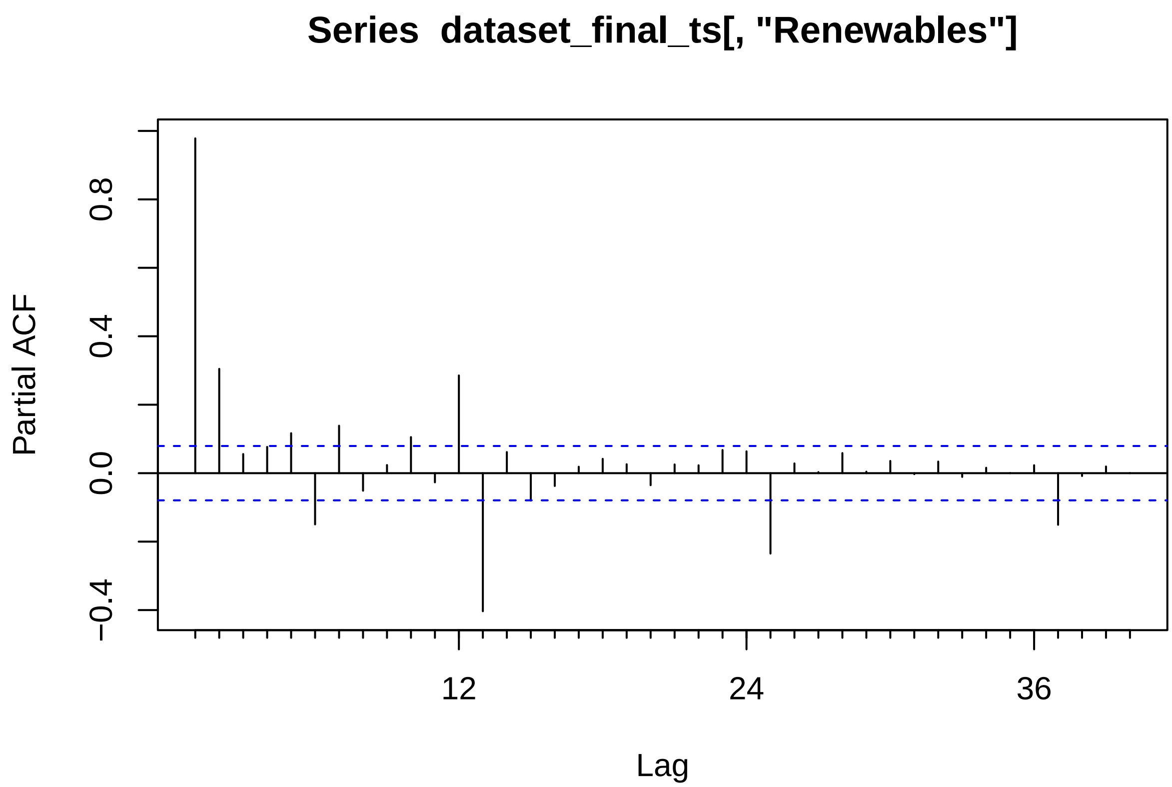
Question 7

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

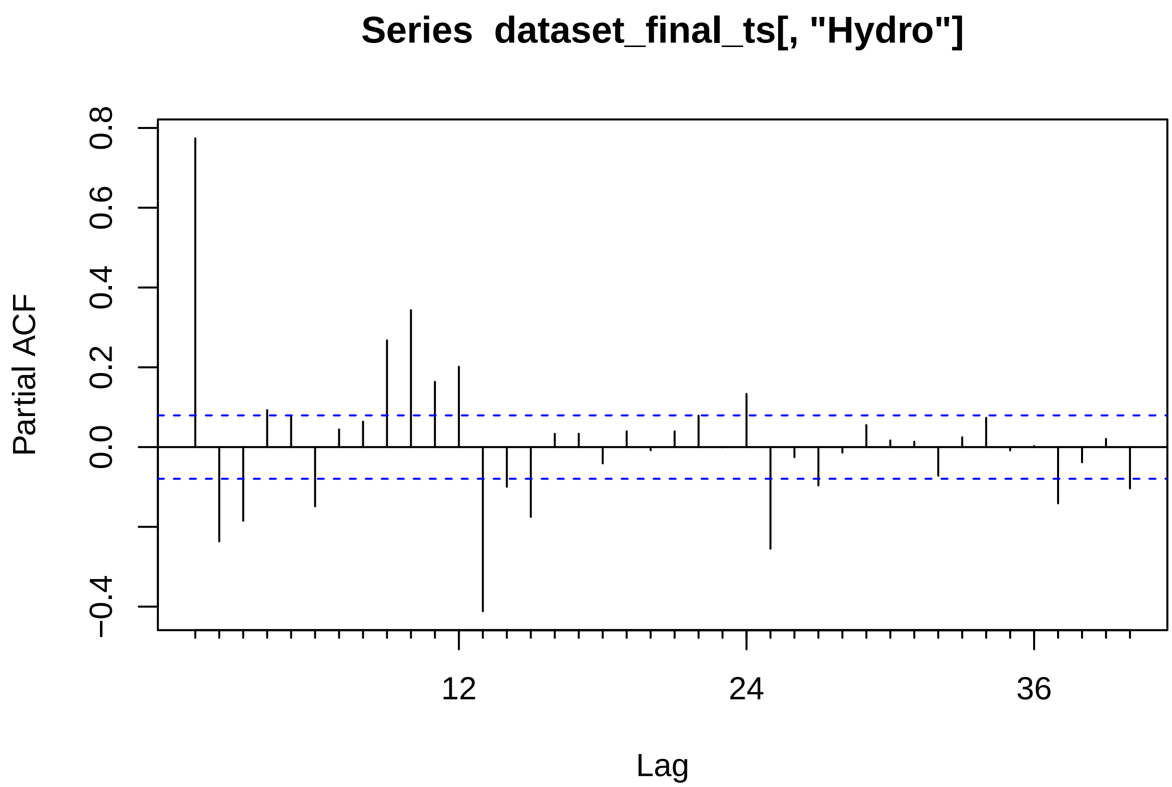
```
Pacf(dataset_final_ts[,"Biomass"], lag.max = 40)
```




```
Pacf(dataset_final_ts[,"Renewables"], lag.max = 40)
```




```
Pacf(dataset_final_ts[,"Hydro"], lag.max = 40)
```




```
#These plots differ significantly from the ACF plots. Not only does the
#autocorrelation fall significantly - reaching almost insignificant levels,
#but the shape of the curve changes as well. In Biomass and RE the shape changes
#from high, positive, mildly declining values to short, insignificant, positive
#and negative values with no trend over time. In hydro, the cyclical trend is
#sort of maintained but the shape is less coherent and the values are much
#closer to 0, especially as the number of lags increases.
```