ENV 790.30 - Time Series Analysis for Energy Data | Spring 2021

Assignment 2 - Due date 01/27/21

Marie McNamara

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 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. Rename the pdf file such that it includes your first and last name (e.g., "LuanaLima_TSA_A02_Sp21.Rmd"). Submit this pdf using Sakai.

```
knitr::opts_chunk$set(echo = TRUE)
#Print out all of the code and the out put of code
```

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.\ install

```
#install.packages("tseries")
#install.packages("dplyr")
#install.packages("forecast")
#install.packages("lubridate")
#install.packages("ggplot2")
#intstall.packages("stats")
#library("xlsx")

library(tseries)
```

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

```
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(forecast)
library(readxl)
library(lubridate)
```

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

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

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 January 2021 Monthly Energy Review. The spreadsheet is ready to be used. Use the command *read.table*() to import the data in R or *panda.read_excel*() in Python (note that you will need to import pandas package). }

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.

```
Table_new <- read.csv("./Data/table102.csv", header=TRUE)
energy <- Table_new[,c(1,4,5,6)]
head(energy)
```

```
##           Month Total.Biomass.Energy.Production
## 1  1973 January                         129.787
## 2 1973 February                         117.338
## 3    1973 March                         129.938
## 4    1973 April                         125.636
## 5      1973 May                         129.834
## 6     1973 June                         125.611
##   Total.Renewable.Energy.Production Hydroelectric.Power.Consumption
## 1                           403.981                         272.703
## 2                           360.900                         242.199
## 3                           400.161                         268.810
## 4                           380.470                         253.185
## 5                           392.141                         260.770
## 6                           377.232                         249.859
```

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().

```
ts_energy <- ts(energy[,2:4])
```

Question 3

Compute mean and standard deviation for these three series.

```
mean(ts_energy[,"Total.Biomass.Energy.Production"])
```

```
## [1] 270.6961
```

```
sd(ts_energy[,"Total.Biomass.Energy.Production"])
```

```
## [1] 87.36311
```

```
mean(ts_energy[,"Total.Renewable.Energy.Production"])
```

```
## [1] 572.7321
```

```
sd(ts_energy[,"Total.Renewable.Energy.Production"])
```

```
## [1] 168.4588
```

```
mean(ts_energy[,"Hydroelectric.Power.Consumption"])
```

```
## [1] 236.9515
```

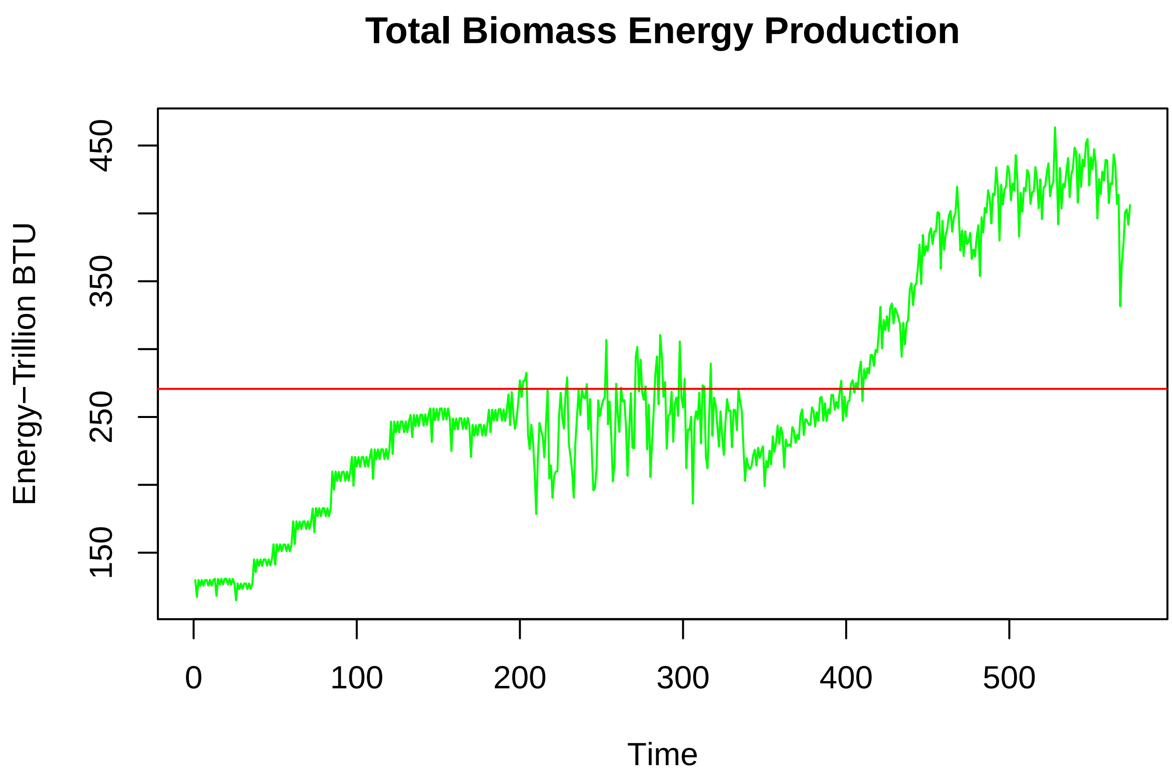
```
sd(ts_energy[,"Hydroelectric.Power.Consumption"])
```

```
## [1] 43.90392
```

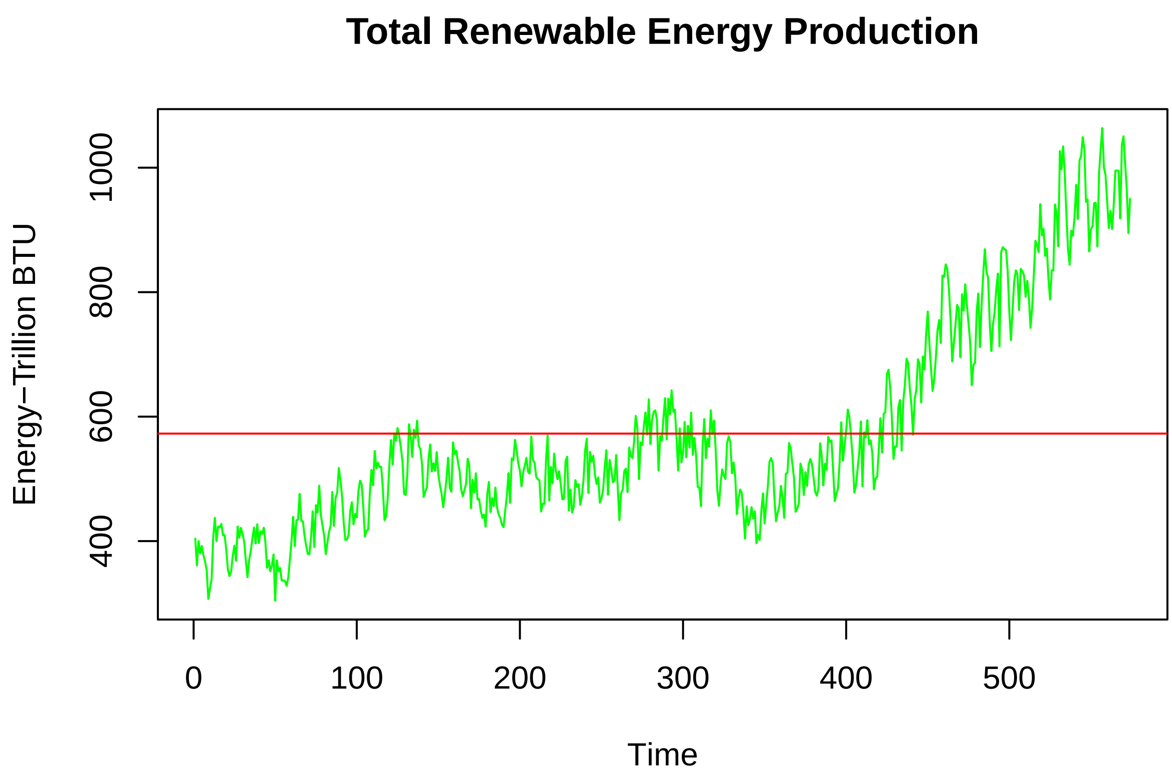
Question 4

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.

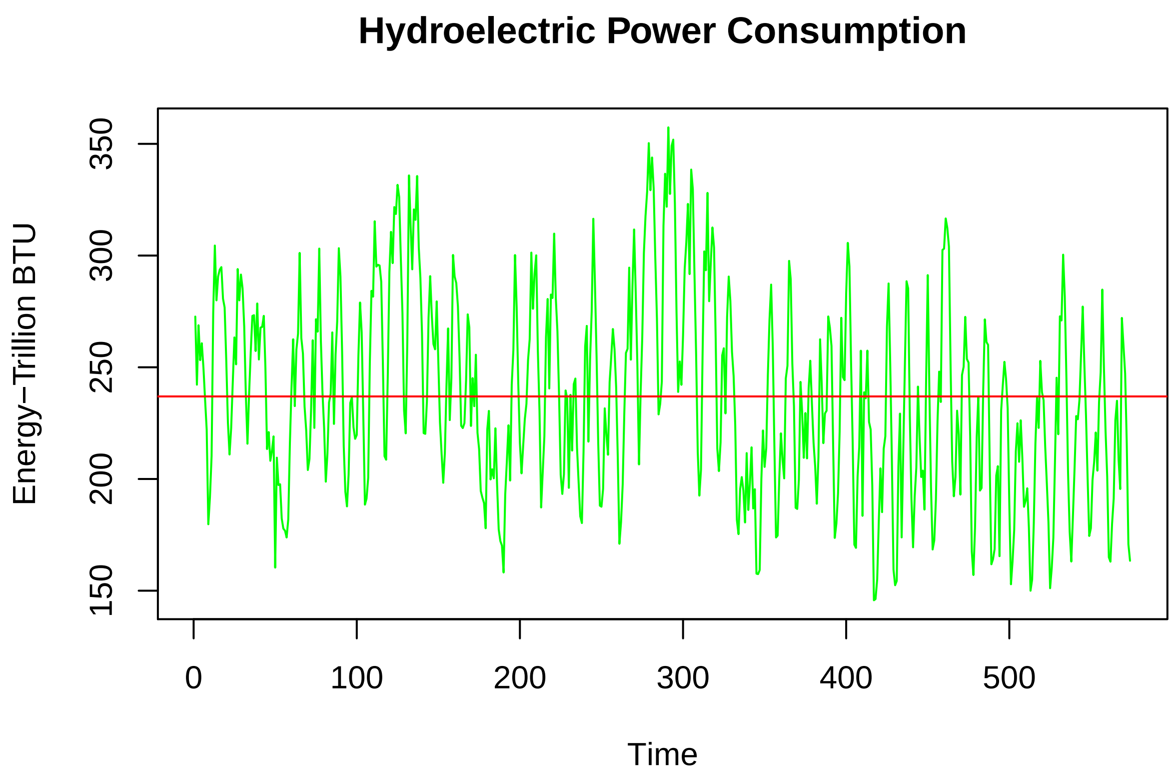
```
plot(ts_energy[,"Total.Biomass.Energy.Production"],type="l",col="green",ylab="Energy-Trillion BTU", xlab
abline(h=mean(ts_energy[,"Total.Biomass.Energy.Production"]), col="red")
```




```
plot(ts_energy[,"Total.Renewable.Energy.Production"],type="l",col="green",ylab="Energy-Trillion BTU", xl
abline(h=mean(ts_energy[,"Total.Renewable.Energy.Production"]), col="red")
```




```
plot(ts_energy[,"Hydroelectric.Power.Consumption"],type="l",col="green",ylab="Energy-Trillion BTU", xlab
abline(h=mean(ts_energy[,"Hydroelectric.Power.Consumption"]), col="red")
```

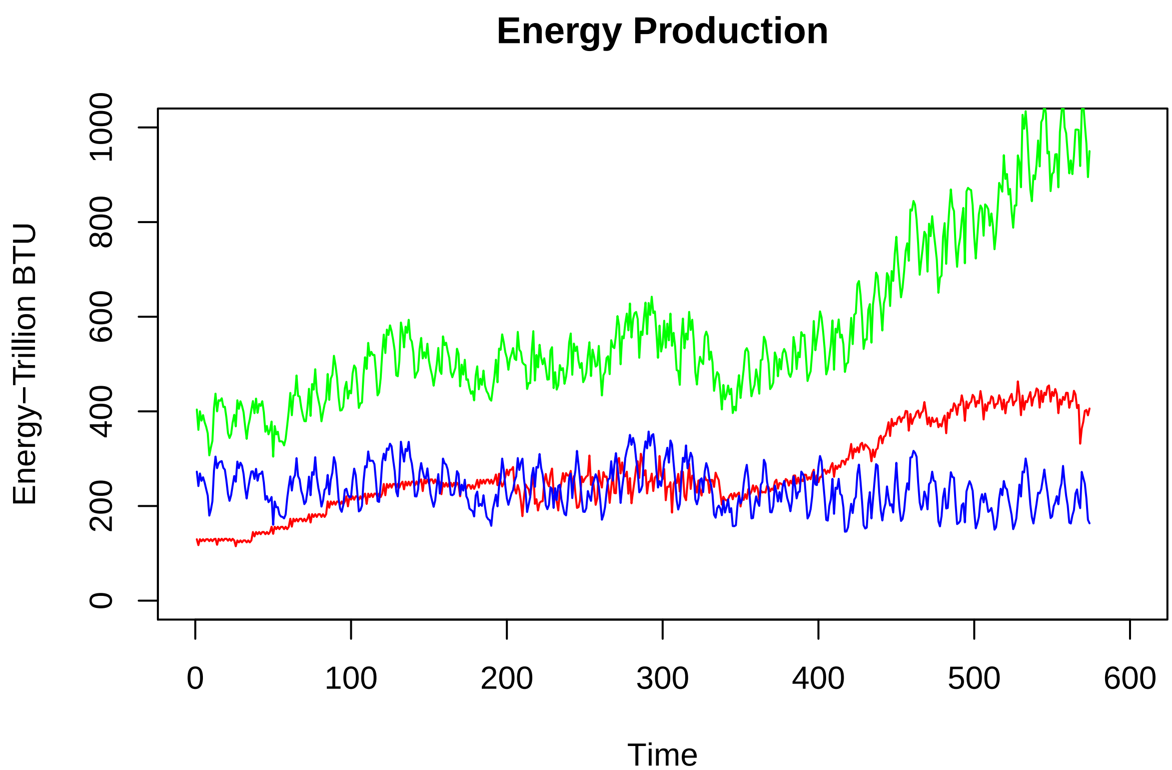


Question 5

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

Below is a graph of all three variables. Based on the graph it is clear that biomass and renewable energy production follow similar tends. They general experience growth and stagnation at the same type. Hydro power production is pretty consistent through time. Based on the calculated correlation coefficients the correlation between biomass and renewable energy is greater than 0.7 and is a significant correlation. The correlation between renewable and hydro power production and biomass energy production are also both less than 0.7 and are not significant.

```
plot(ts_energy[,"Total.Biomass.Energy.Production"],type="l",col="red",ylab="Energy-Trillion BTU", xlab=
lines(ts_energy[,"Total.Renewable.Energy.Production"], col =  "green")
lines(ts_energy[,"Hydroelectric.Power.Consumption"], col =  "blue")
```




```
Bio <- ts_energy[,c(1)]
renew <- ts_energy[,c(2)]
hydro <- ts_energy[,c(3)]
cor(Bio,renew)
```

```
## [1] 0.9234609
```

```
cor(renew,hydro)
```

```
## [1] -0.002756852
```

```
cor(hydro,Bio)
```

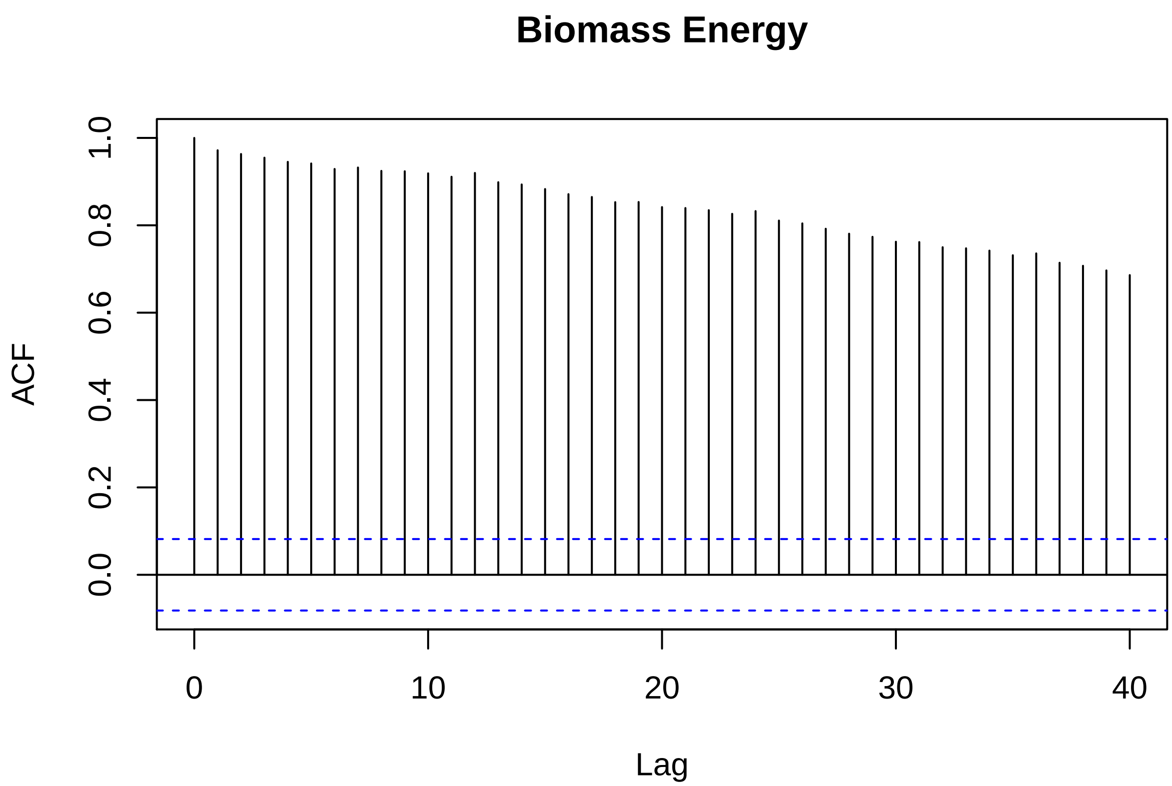
```
## [1] -0.2555675
```

Question 6

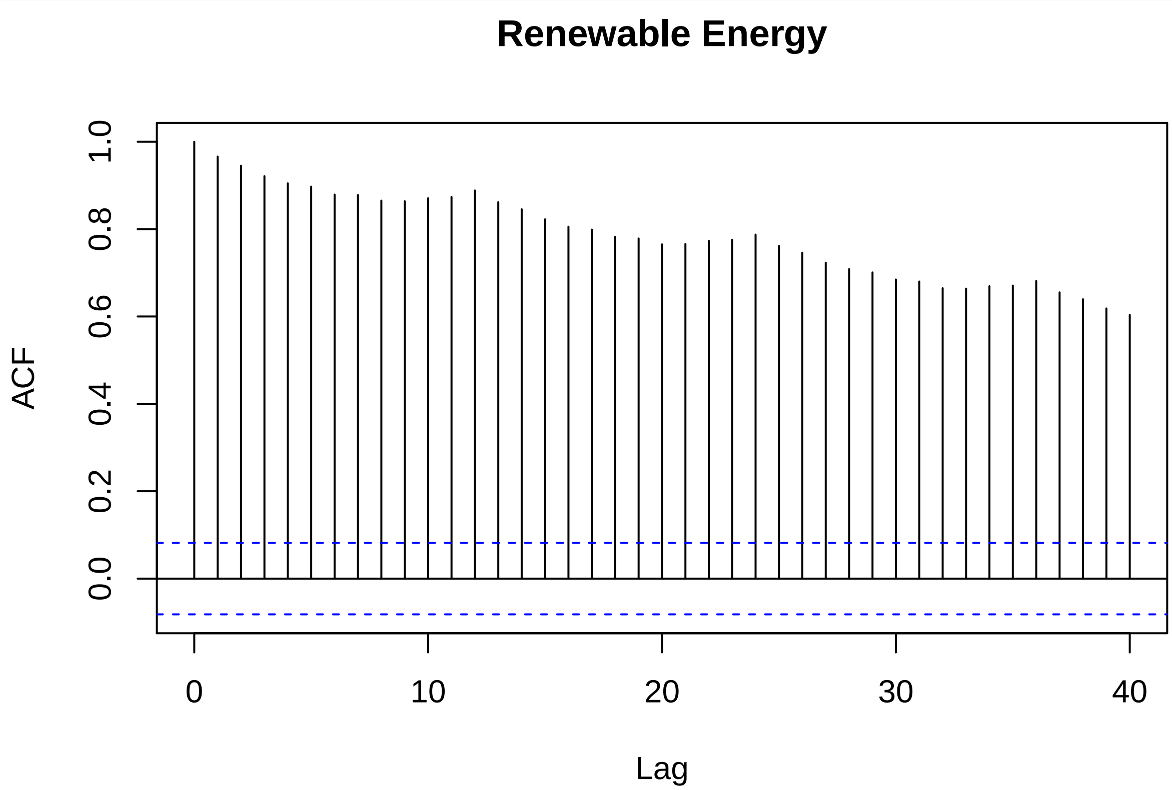
Compute the autocorrelation function from lag 1 up to lag 40 for these three variables. What can you say about these plots? Do the three of them have the same behavior?

Hydropower has clear seasonal variation with the fluctuations in lags rising and falling. This plot is rather different from the renewable energy ACF and biomass energy ACF. There is a strong correlation with initial lags on the renewable energy plot. This is similar to biomass energy ACF plot, again there is a strong correlation between initial lags in from biomass energy. The correlation then gets weaker as the lags progress.

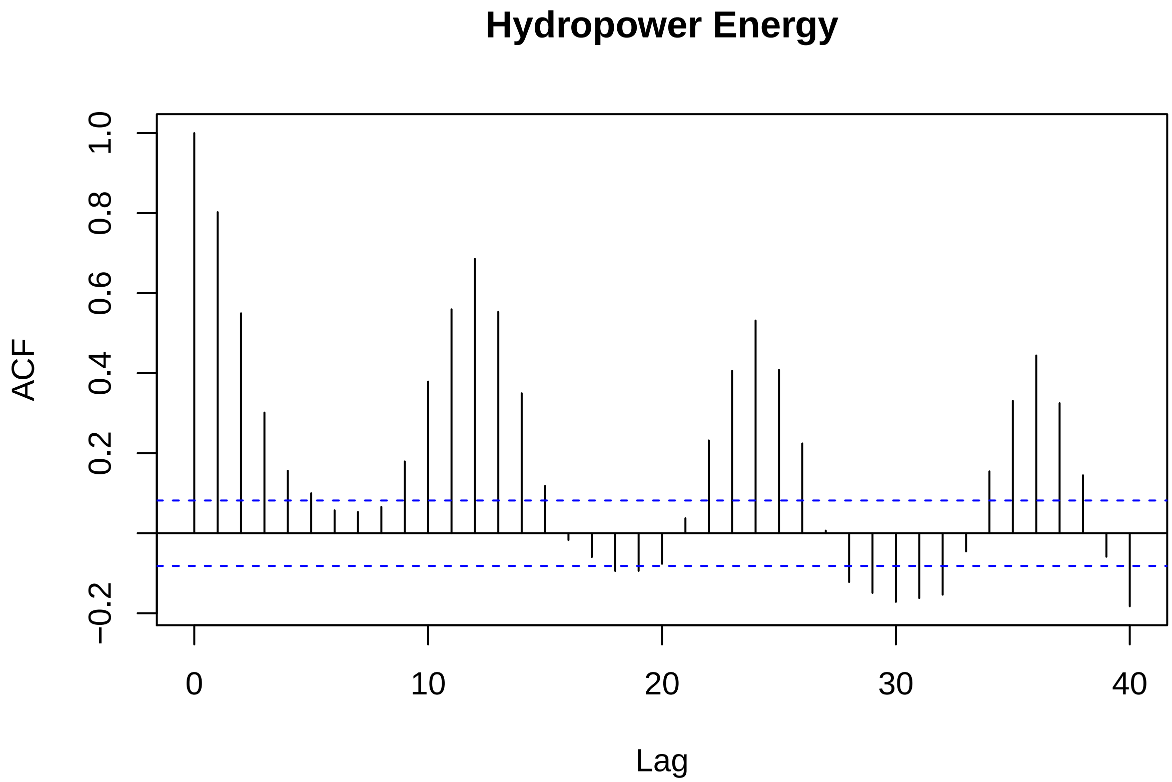
```
ts_energy_acf_1=acf(ts_energy[,1],lag.max=40, type="correlation", plot = TRUE, main="Biomass Energy")
```




```
ts_energy_acf_2=acf(ts_energy[,2],lag.max=40, type="correlation", plot = TRUE, main="Renewable Energy")
```




```
ts_energy_acf_3=acf(ts_energy[,3],lag.max=40, type="correlation", plot = TRUE, main="Hydropower Energy")
```

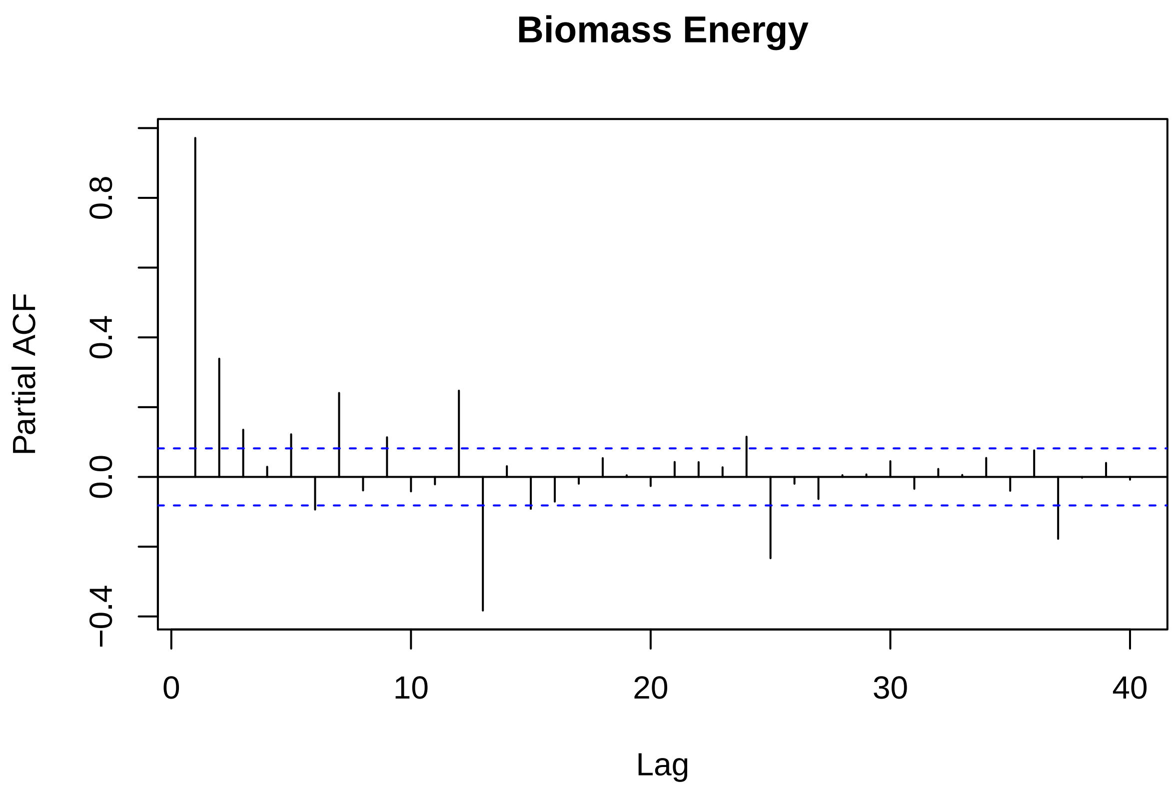


Question 7

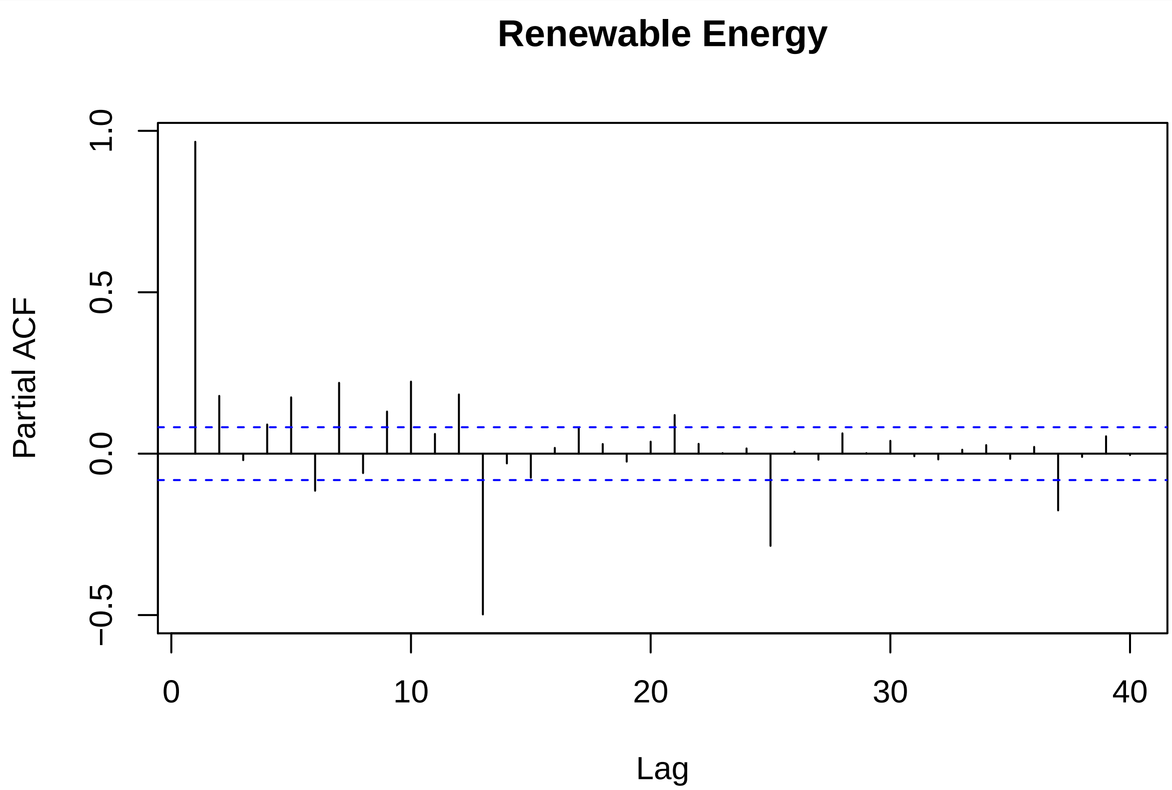
Compute the partial autocorrelation function from lag 1 to lag 40 for these three variables. How these plots differ from the ones in Q6?

The correlation between lags dramatically falls for the PACF chart once the intermediate variables are removed for all three charts from the ACF charts. The PACF charts for the renewable Energy and biomass energy production correlation between lags is dramatically less significant than the results obtained from the ACF charts. There are some interesting correlations produced in the PACF charts, the correlation coefficients in the biomass and renewable energy PACF charts are positive but there are a few negative correlations. For Hydro-power production the seasonal variations seen in the ACF chart do not occur consistently in the PACF chart and the correlations are less profound. For hydro-power the lag correlations from 0-14 are more significant than the lags from 14 onward.

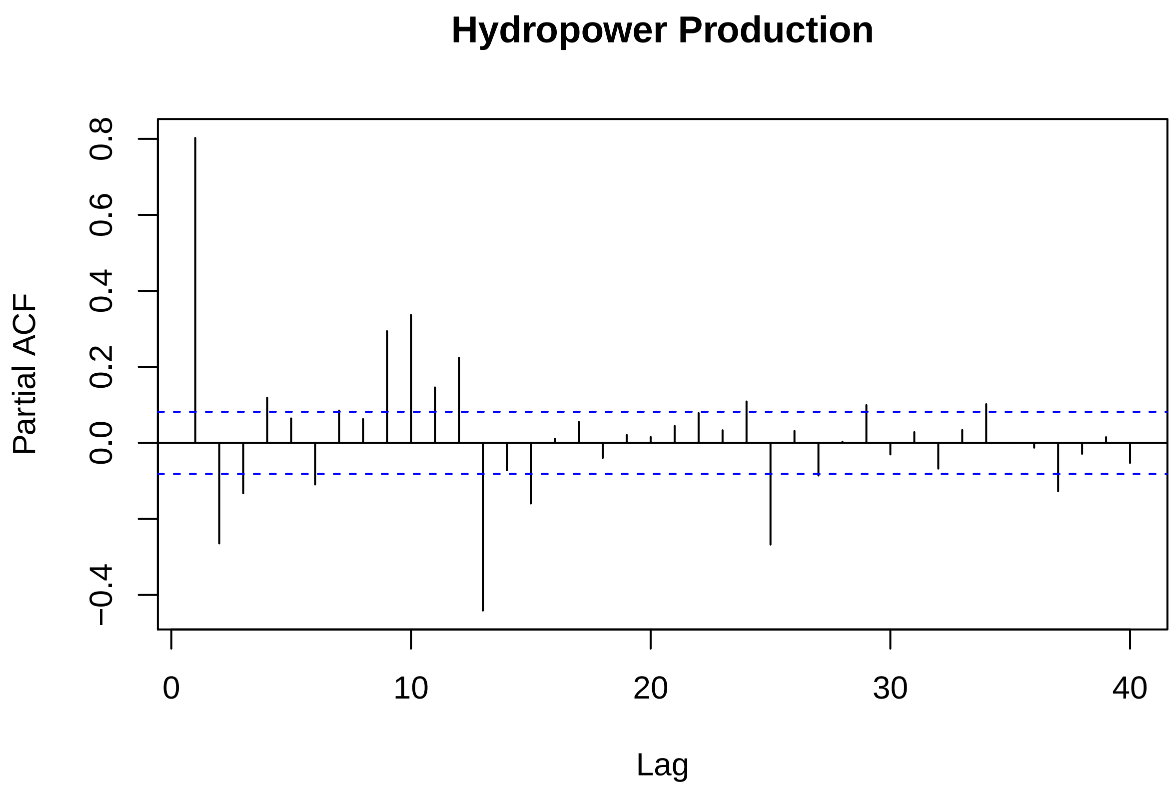
```
ts_pacf_1=pacf(ts_energy[,1],lag.max=40, plot = TRUE,main="Biomass Energy")
```




```
ts_pacf_2=pacf(ts_energy[,2],lag.max=40, plot = TRUE,main="Renewable Energy")
```




```
ts_pacf_2=pacf(ts_energy[,3],lag.max=40, plot = TRUE,main="Hydropower Production")
```


Hydropower Production
Partial ACF
0.8
0.6
0.4
0.2
0.0
−0.4
0
10
20
30
40
Lag