ENV 790.30 - Time Series Analysis for Energy Data | Spring 2023

Assignment 2 - Due date 02/03/23

Maggie O'Shea

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_Sp23.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

#Used '#' to make this code available for reproducibility but will not run every time I use the script.
#install.packages("forecast")
#install.packages("tseries")
#install.packages("dplyr")

library(forecast)
library(tseries)
library(dplyr)
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 2022 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*.

```r
#Importing data set
renewableenergy <- read.table(file=
                    "../Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source-Edit.csv",
                              header = TRUE,
                              sep=",",
                              stringsAsFactors=TRUE)
```

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
clean_RE <- renewableenergy[,c(1, 4:6)]
head(clean_RE)
```

```
##            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().

```r
ts_renewables <- ts(clean_RE[,2:4], frequency = 12, start = c(1973, 1))
head(ts_renewables)
```

```
##          Total.Biomass.Energy.Production Total.Renewable.Energy.Production
## Jan 1973                         129.787                           403.981
## Feb 1973                         117.338                           360.900
## Mar 1973                         129.938                           400.161
## Apr 1973                         125.636                           380.470
## May 1973                         129.834                           392.141
## Jun 1973                         125.611                           377.232
##          Hydroelectric.Power.Consumption
## Jan 1973                         272.703
## Feb 1973                         242.199
## Mar 1973                         268.810
## Apr 1973                         253.185
```

```
## May 1973                           260.770
## Jun 1973                           249.859
```

Question 3

Compute mean and standard deviation for these three series.

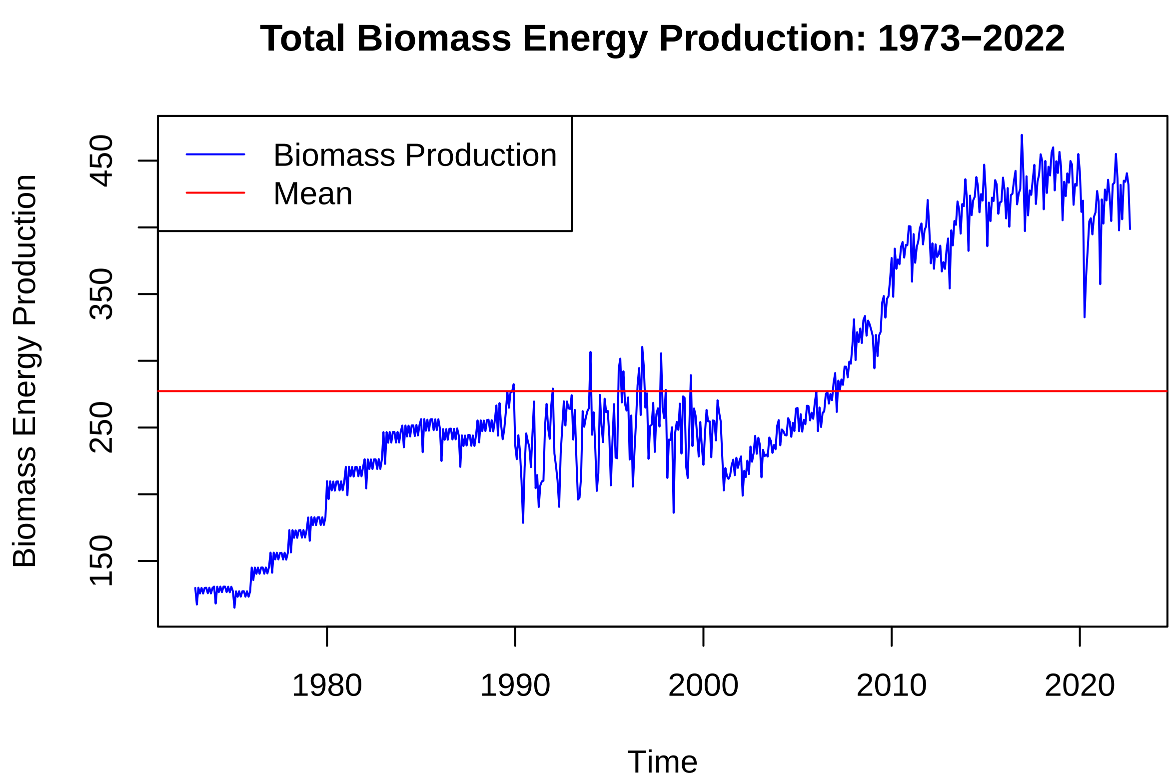
```
mean_biomass <- mean(clean_RE$Total.Biomass.Energy.Production)
mean_total <- mean(clean_RE$Total.Renewable.Energy.Production)
mean_hydro <- mean(clean_RE$Hydroelectric.Power.Consumption)

sd_biomass <- sd(clean_RE$Total.Biomass.Energy.Production)
sd_total <- sd(clean_RE$Total.Renewable.Energy.Production)
sd_hydro <- sd(clean_RE$Hydroelectric.Power.Consumption)
```

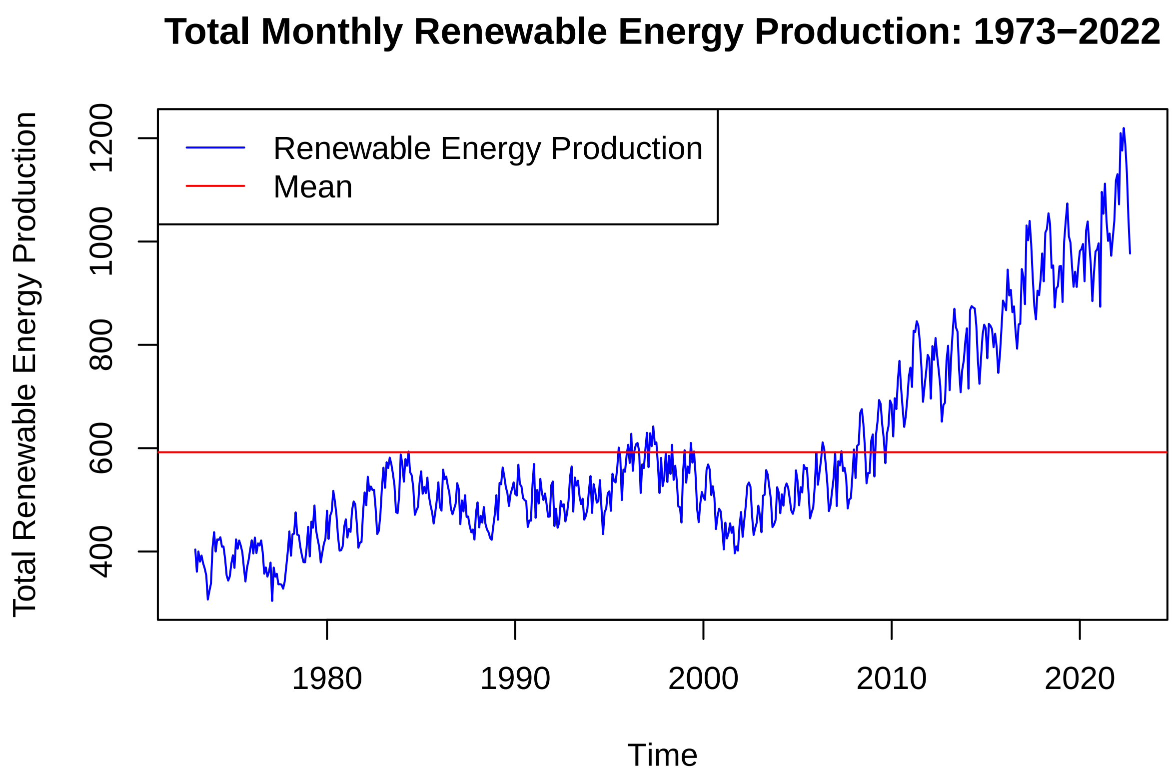
Question 4

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.

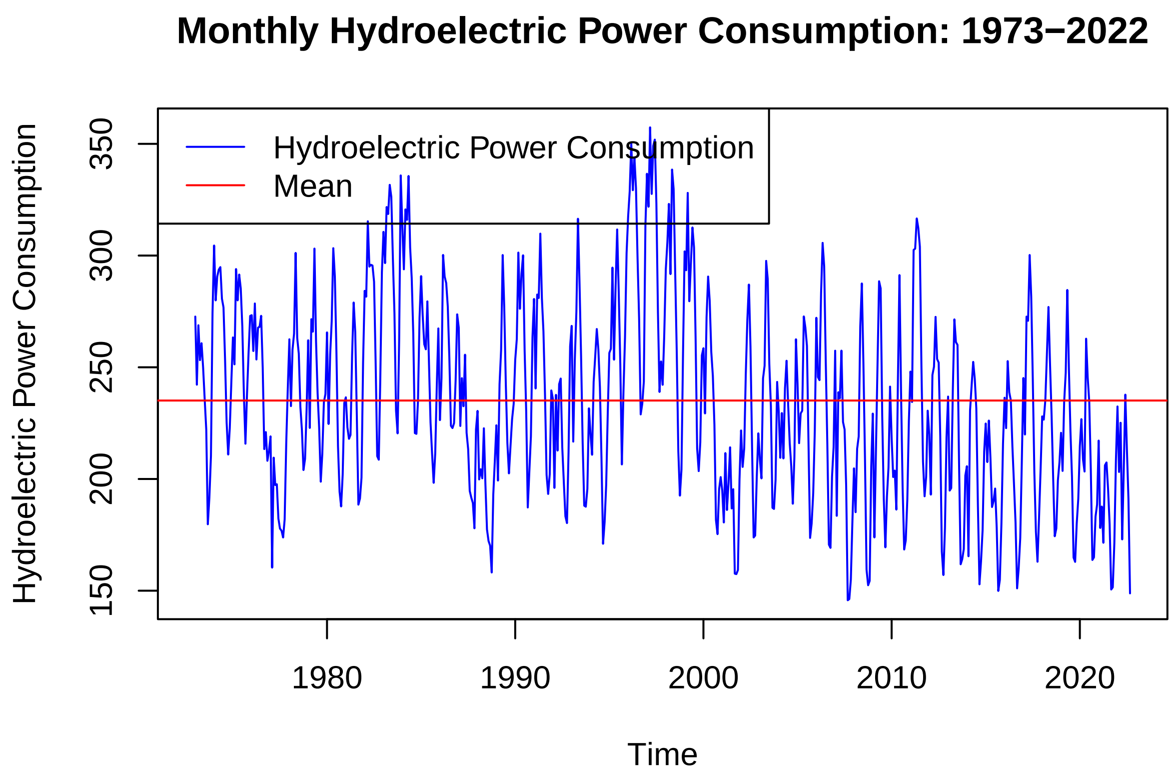
```
plot(ts_renewables[,"Total.Biomass.Energy.Production"],type="l",col="blue",
     ylab="Biomass Energy Production",xlab="Time",
     main="Total Biomass Energy Production: 1973-2022")
abline(h=mean(ts_renewables[,"Total.Biomass.Energy.Production"]),col="red")
legend("topleft",legend=c("Biomass Production","Mean"),
       lty=c("solid","solid"),col=c("blue","red"))
```



```r
plot(ts_renewables[,"Total.Renewable.Energy.Production"],type="l",col="blue",
     ylab="Total Renewable Energy Production",xlab="Time",
     main="Total Monthly Renewable Energy Production: 1973-2022")
abline(h=mean(ts_renewables[,"Total.Renewable.Energy.Production"]),col="red")
legend("topleft",legend=c("Renewable Energy Production","Mean"),
       lty=c("solid","solid"),col=c("blue","red"))
```




```r
plot(ts_renewables[,"Hydroelectric.Power.Consumption"],type="l",col="blue",
     ylab="Hydroelectric Power Consumption",xlab="Time",
     main="Monthly Hydroelectric Power Consumption: 1973-2022")
abline(h=mean(ts_renewables[,"Hydroelectric.Power.Consumption"]),col="red")
legend("topleft",legend=c("Hydroelectric Power Consumption","Mean"),
       lty=c("solid","solid"),col=c("blue","red"))
```



Question 5

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

```
cor(clean_RE$Total.Biomass.Energy.Production, clean_RE$Total.Renewable.Energy.Production)
```

```
## [1] 0.9185941
```

```
cor(clean_RE$Total.Biomass.Energy.Production, clean_RE$Hydroelectric.Power.Consumption)
```

```
## [1] -0.2998201
```

```
cor(clean_RE$Total.Renewable.Energy.Production, clean_RE$Hydroelectric.Power.Consumption)
```

```
## [1] -0.09958758
```

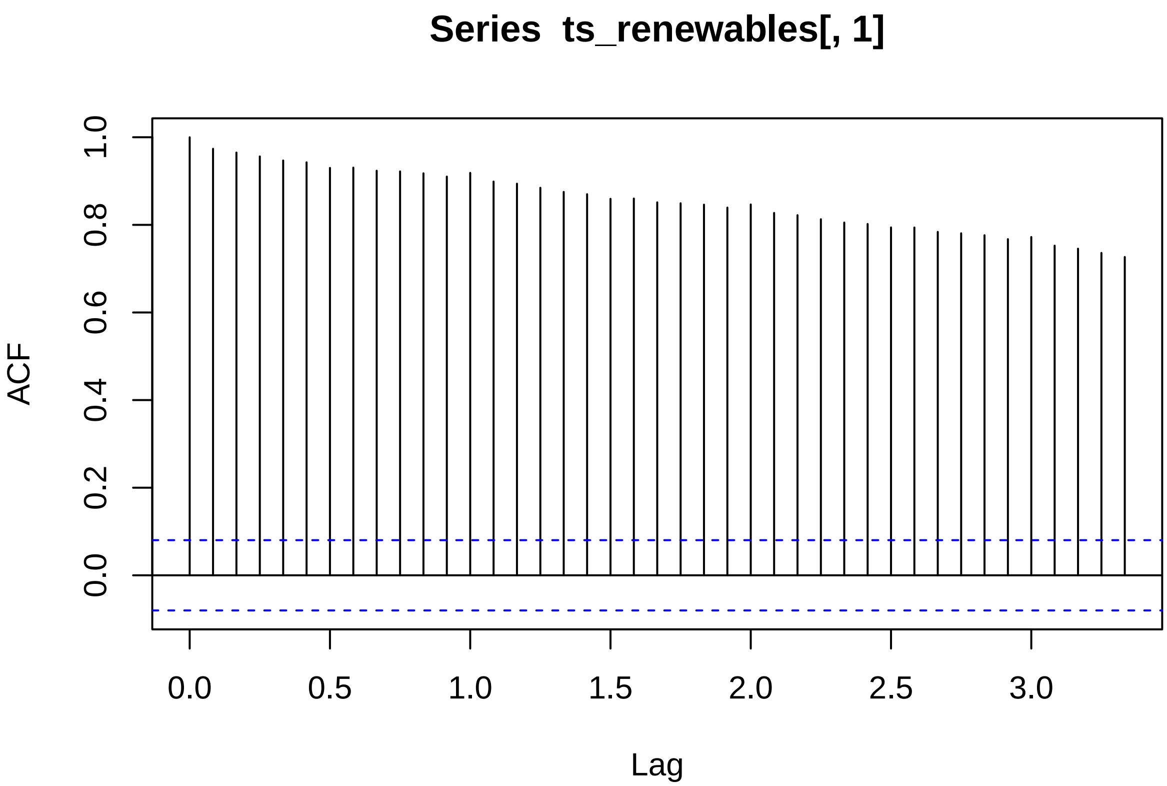
Question 5 Answer:

Total Biomass Energy Production and Total Renewable Energy Production are significantly correlated (correlation: ~0.92). Looking at the graphs in question 4, this is unsurprising as they both seem to increase in similar ways and of course, biomass production is likely an input or one piece of the total renewable production number. The other correlations are not as strong – the second strongest is between biomass energy production and hydroelectric power with a correlation of -0.2998. This indicates that when one increases the other may decrease – this conceptually is possible because it is possible that when biomass production increases then the demand or consumption for hydroelectric power decreases because an alternative is available, that being biomass. Finally, Total Renewable Energy Production and Hydroelectric Consumption is not very correlated (correlation: -0.0996). I would have expected a slightly stronger relationship given that hydroelectric power consumption is of course related to hydroelectric production which would be one part of the total renewable energy production, however because it is a distant relationship it is not entirely unexpected that they would not be more strongly correlated given all the other inputs to total renewable energy production.

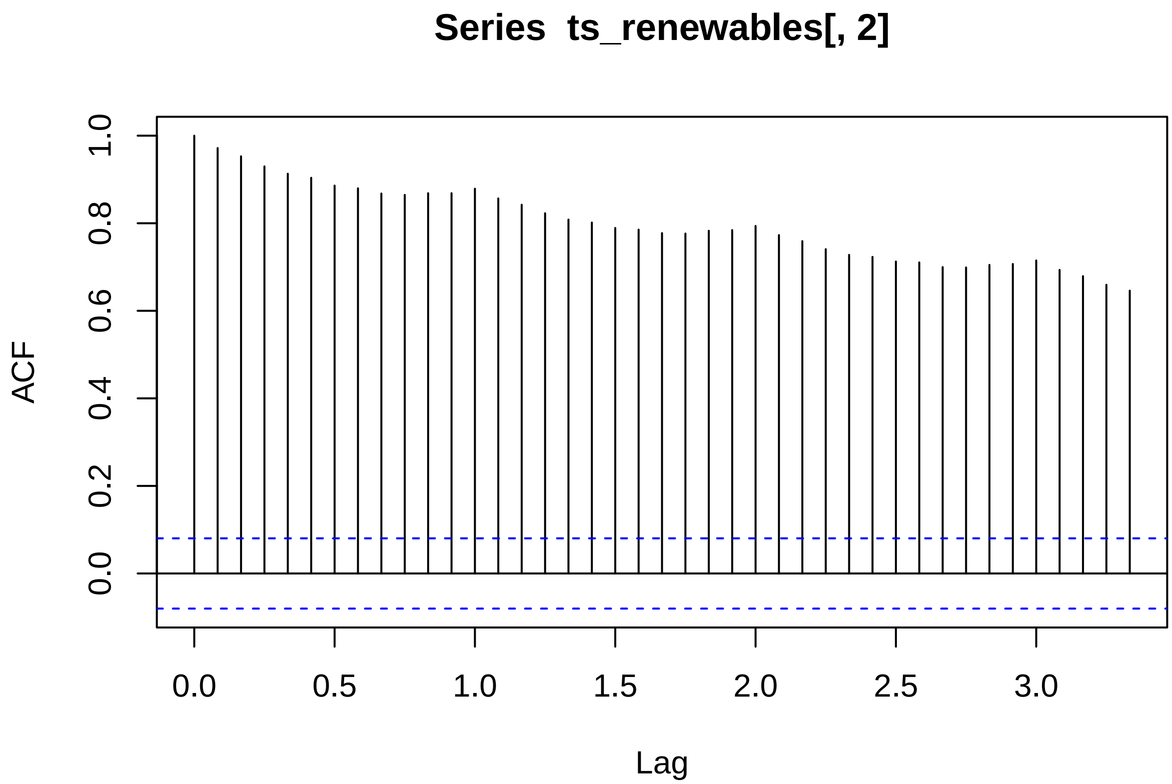
Question 6

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?

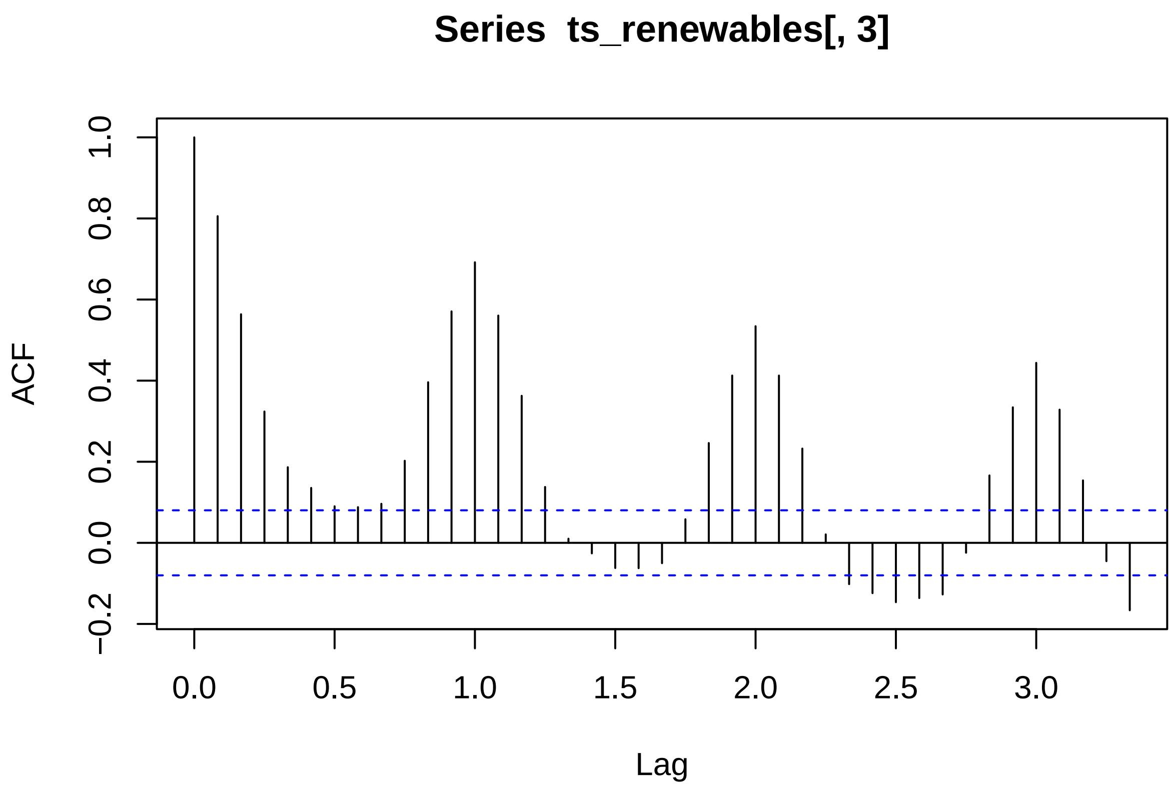
```r
acf(ts_renewables[,1], lag.max = 40)
```




```r
acf(ts_renewables[,2], lag.max = 40)
```




```
acf(ts_renewables[,3], lag.max = 40)
```



###

Question 6 Answer:

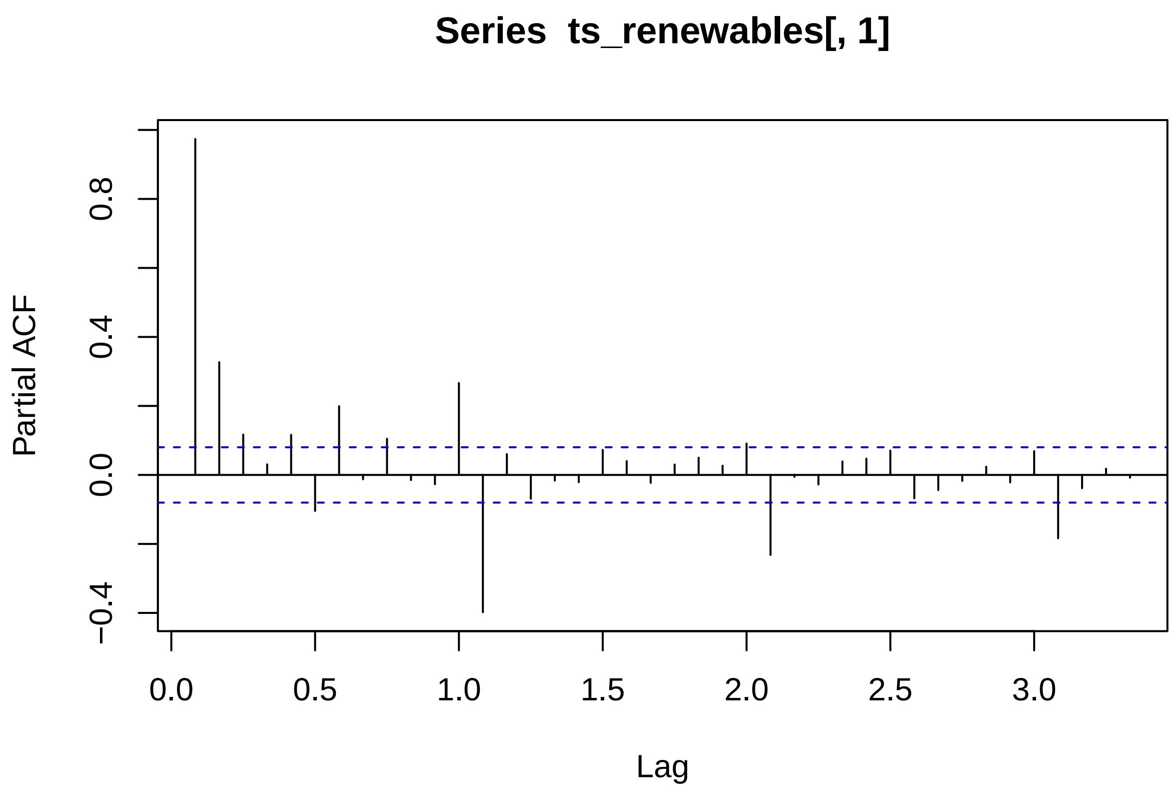
These three plots do not all have the same behavior. Plots 1 and 2, which represent Biomass Production and

Total Renewable Energy Production are similar in that they have high positive correlation even towards the end of the plot after many lags. The highest correlations in both are 1, and the lowest for Total Renewable Energy is ~0.7 and for Biomass production ~0.75. They both seem to show a similar declining trend. Plot 2 differs slightly from plot 1 because it has a slight seasonality to it - though the correlations do decline similar to plot 1, they also have a slight increase and decrease that indicates seasonality. Finally, plot 3 shows definite seasonality. It appears that correlations are very strong at lag 0.0 (correlation 1.0) but decline over time and even move to negative numbers. However, for all of these plots, correlation even at the final lag are significant beyond the blue line (though less so for the final plot).

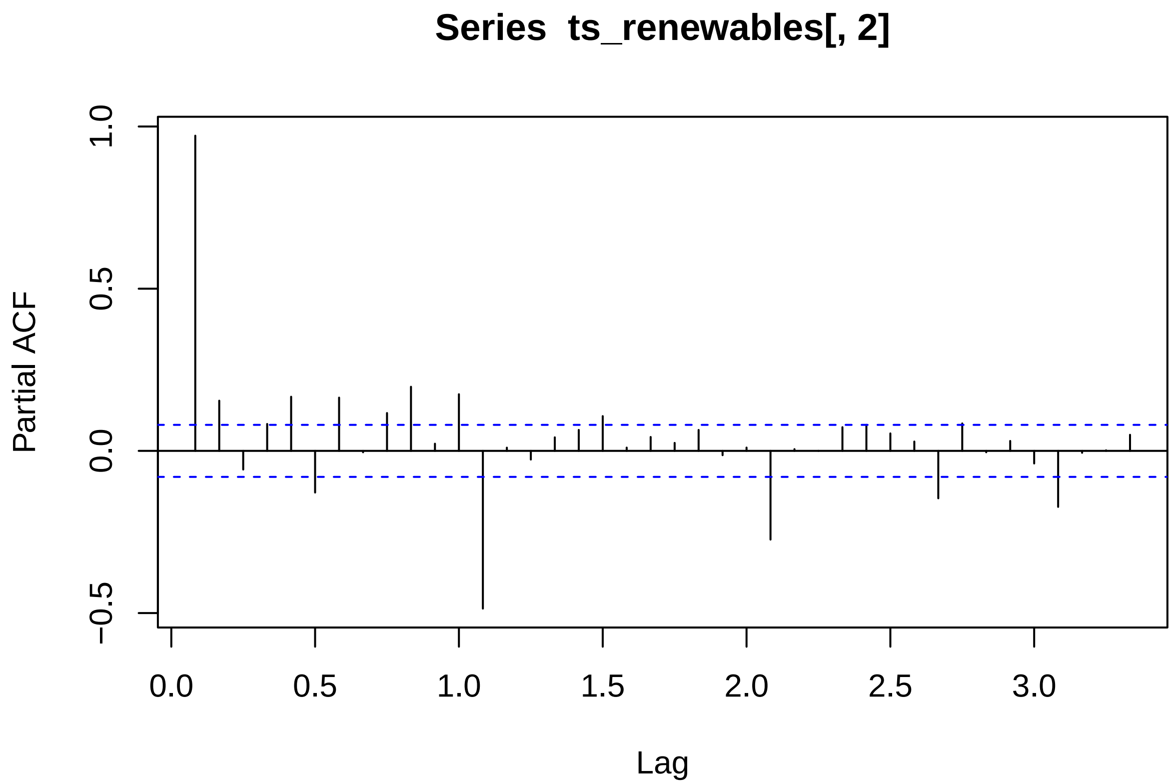
Question 7

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

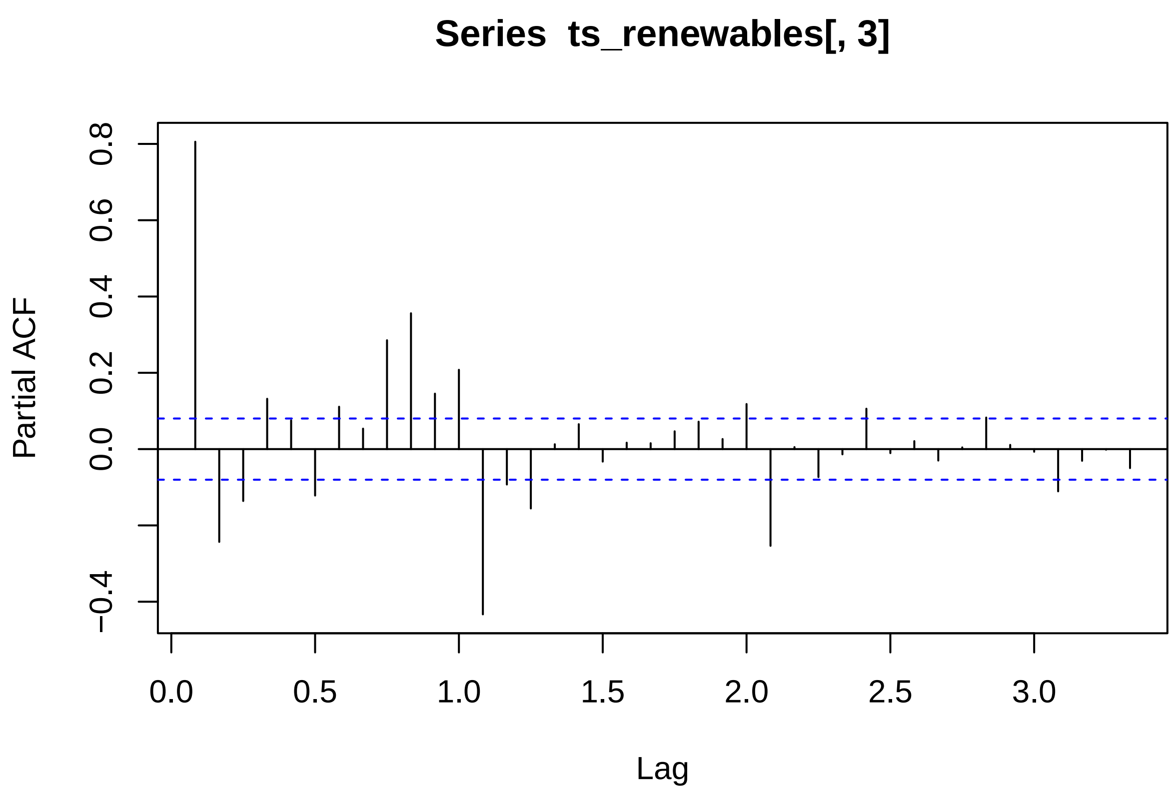
```
pacf(ts_renewables[,1], lag.max = 40)
```




```
pacf(ts_renewables[,2], lag.max = 40)
```




```
pacf(ts_renewables[,3], lag.max = 40)
```



###

Question 7 Answer:

These plots are very different from the ACF plots. While in the first ACF plots showed high positive (though declining) correlations for plots biomass and total renewables, the PACF for these instead show both positive and negative correlations that are less strong than in the previous plot. The final plot, hydroelectric power consumption, maintains some of the seasonality in the early stage of the plot (lag 0-1.5) but ultimately the correlations become quite weak after lag 1.5 and especially by 2.0.