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

Assignment 3 - Due date 02/01/24

Julia Kagiliery

Directions

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).

Please keep this R code chunk options for the report. It is easier for us to grade when we can see code and output together. And the tidy.opts will make sure that line breaks on your code chunks are automatically added for better visualization.

When you have completed the assignment, **Knit** the text and code into a single PDF file. Submit this pdf using Sakai.

Questions

Consider the same data you used for A2 from the spreadsheet "Table_10.1_Renewable_Energy_Production_and_Consumption The data comes from the US Energy Information and Administration and corresponds to the December 2022 **Monthly** Energy Review. Once again 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.

R packages needed for this assignment:"forecast","tseries", and "Kendall". 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(Kendall)
library(ggplot2)
library(lubridate)
```

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

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

```
library(readxl)
library(cowplot)
```

```
##
## Attaching package: 'cowplot'

## The following object is masked from 'package:lubridate':
##
##     stamp
```

```
DataPandC <-
  read_excel(
    "~/Julia_Kagiliery_TSA_Sp24/Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source.xl
    skip = 10,
    sheet = "Monthly Data",
    col_names = TRUE
  )
```

```
DataPandC <- DataPandC[,c(1,4:6)]
DataPandC <- DataPandC[-1, , drop = FALSE]
```

##Trend Component

Q1

For each time series, i.e., Renewable Energy Production and Hydroelectric Consumption create three plots: one with time series, one with the ACF and with the PACF. You may use the some code form A2, but I want all the three plots side by side as in a grid. (Hint: use function `plot_grid()` from the `cowplot` package)

#Set Up:

```
year1 <- year(DataPandC$Month[1])
month1 <- month(DataPandC$Month[1])
```

```
DataPandCTS <- DataPandC[,-1] |>
    ts(start = c(year1,month1), frequency = 12, deltat = 1/12)
```

#Biomass Enegry:

inspiration for plot = FLASE from ChatGPT. Prompt: "How do I use plot_grid() with PCF and ACF in R"

```
P1ACF <-
  Acf(
    DataPandCTS[, 1],
    lag.max = 40,
    main = paste("Biomass Production", label_size = 3),
    plot = FALSE
  )

P2PCF <-
  Pacf(
    DataPandCTS[, 1],
    lag.max = 40,
    main = paste("Biomass Production", label_size = 3),
```

```
    plot = FALSE
  )
```

```
P1 <- autoplot(P1ACF)
P2 <- autoplot(P2PCF)
P3 <- DataPandCTS[, 1] |>
  autoplot(color = "darkblue") +
  ylab("Biomass Production") +
  xlab("Year")
```

#Renewable Energy:

```
P4ACF <-
  Acf(
    DataPandCTS[, 2],
    lag.max = 40,
    main = paste("Renewable Production", label_size = 3),
    plot = FALSE
  )

P5PCF <-
  Pacf(
    DataPandCTS[, 2],
    lag.max = 40,
    main = paste("Renewable Production", label_size = 3),
    plot = FALSE
  )
```

```
P4 <- autoplot(P4ACF)
P5 <- autoplot(P5PCF)
P6 <- DataPandCTS[, 2] |>
  autoplot(color = "blue", label_size = 3) +
  ylab("Renewable Production") +
  xlab("Year")
```

```
## Warning in ggplot2::geom_line(na.rm = TRUE, ...): Ignoring unknown parameters:
## `label_size`
```

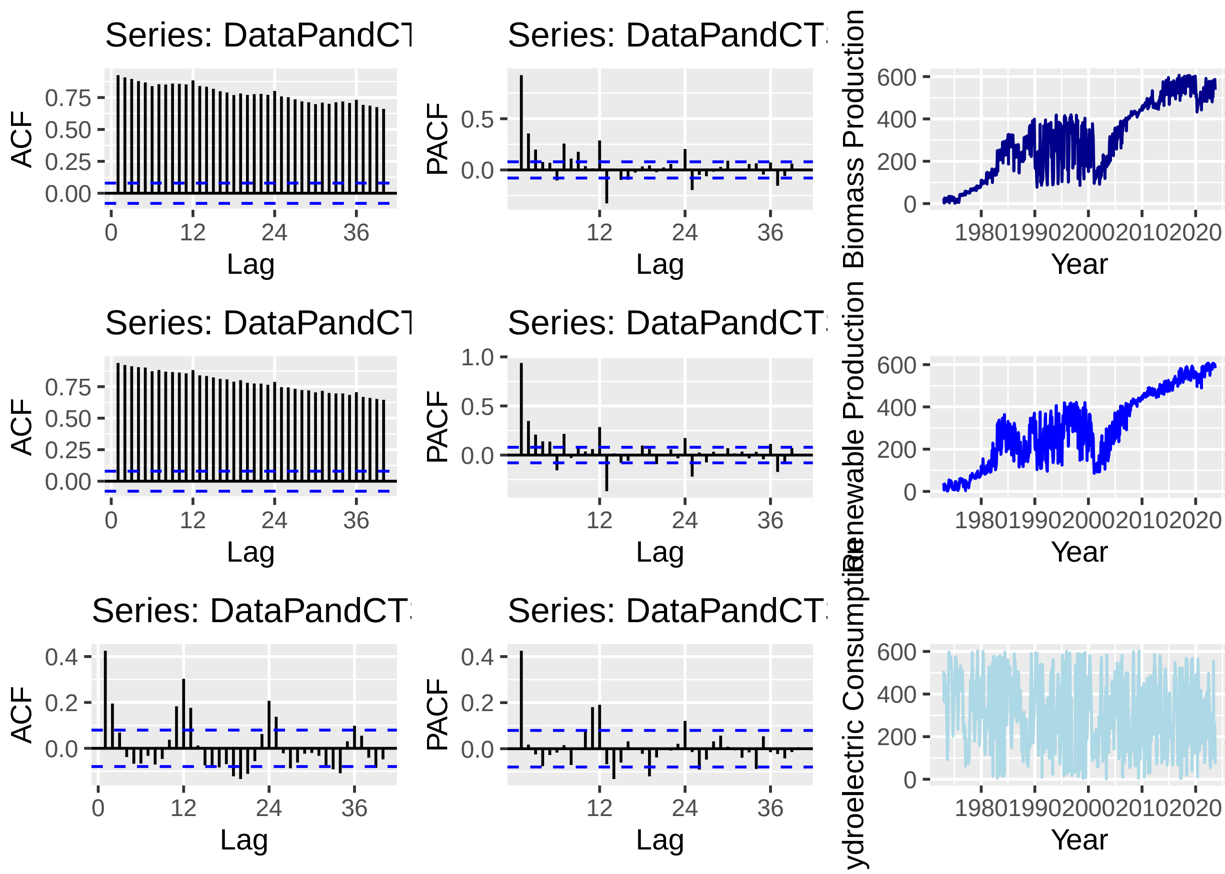
#Hydroelectric Energy:

```
P7ACF <-
  Acf(
    DataPandCTS[, 3],
    lag.max = 40,
    main = paste("Hydroelectric Consumption", label_size = 3),
    plot = FALSE
  )

P8PCF <-
  Pacf(
    DataPandCTS[, 3],
    lag.max = 40,
    main = paste("Hydroelectric Consumption", label_size = 3),
    plot = FALSE
  )
```

```r
P7 <- autoplot(P7ACF)
P8 <- autoplot(P8PCF)
P9 <- DataPandCTS[, 3] |>
  autoplot(color = "lightblue", label_size = 3) +
  ylab("Hydroelectric Consumption") +
  xlab("Year")
```

```
## Warning in ggplot2::geom_line(na.rm = TRUE, ...): Ignoring unknown parameters:
## `label_size`
```

```r
plot_grid(P1, P2, P3, P4, P5, P6, P7, P8, P9, nrow = 3, label_size = 3)
```



Q2

From the plot in Q1, do the series Total Biomass Energy Production, Total Renewable Energy Production, Hydroelectric Power Consumption appear to have a trend? If yes, what kind of trend?

It appears as though the total biomass energy production and the total renewable energy production each have a positive linear trend in that the production increases over time. However, this trend does not seem to be present in total hydroelectric energy consumption. All three time series seem to be influenced by some random and seasonal effects as well.

Q3

Use the *lm()* function to fit a linear trend to the three time series. Ask R to print the summary of the regression. Interpret the regression output, i.e., slope and intercept. Save the regression coefficients for further analysis.

total number of observations = 609

```r
t <- c(1:609)
```

#Trend for Biomass Energy Production:

```r
DataBiomass <- DataPandC$`Total Biomass Energy Production`

DataBiomass <- ts(DataBiomass)

BiomassTrend = lm(DataBiomass ~ t)

summary(BiomassTrend)
```

```
##
## Call:
## lm(formula = DataBiomass ~ t)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -102.344  -23.754    5.491   31.980   83.154
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 134.27841    3.18601   42.15   <2e-16 ***
## t             0.47713    0.00905   52.72   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 39.26 on 607 degrees of freedom
## Multiple R-squared:  0.8208, Adjusted R-squared:  0.8205
## F-statistic:  2780 on 1 and 607 DF,  p-value: < 2.2e-16
```

The incertcept for this linear trend:

```r
Bbeta0 = as.numeric(BiomassTrend$coefficients[1]) |>
  print()
```

```
## [1] 134.2784
```

The slope for this linear trend:

```r
Bbeta1 = as.numeric(BiomassTrend$coefficients[2]) |>
  print()
```

```
## [1] 0.477135
```

Here, my p-value is less than 0.05 so my coefficient is significant and there is a trend.

#Trend for Renewable Energy Production:

```r
DataRenewable <- DataPandC$`Total Renewable Energy Production`

DataRenewable <- ts(DataRenewable) # should this be a time series?

RenewableTrend = lm(DataRenewable ~ t)

summary(RenewableTrend)
```

```
##
```

```
## Call:
## lm(formula = DataRenewable ~ t)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -148.27  -35.63   11.58   41.51  144.27
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 180.98940    4.90151   36.92   <2e-16 ***
## t             0.70404    0.01392   50.57   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 60.41 on 607 degrees of freedom
## Multiple R-squared:  0.8081, Adjusted R-squared:  0.8078
## F-statistic:  2557 on 1 and 607 DF,  p-value: < 2.2e-16
```

The incertcept for this linear trend:

```r
Rbeta0 = as.numeric(RenewableTrend$coefficients[1]) |>
  print()
```

```
## [1] 180.9894
```

The slope for this linear trend:

```r
Rbeta1 = as.numeric(RenewableTrend$coefficients[2]) |>
  print()
```

```
## [1] 0.7040391
```

Here, my p-value is less than 0.05 so my coefficient is significant and there is a trend.

#Trend for Hyrdoelectric Energy Consumption:

```r
DataHydro <- DataPandC$`Hydroelectric Power Consumption`

DataHydro <- ts(DataHydro)

HydroTrend = lm(DataHydro ~ t)

summary(HydroTrend)
```

```
##
## Call:
## lm(formula = DataHydro ~ t)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -29.818 -10.620  -0.669   9.357  39.528
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 82.734747   1.140265  72.557  < 2e-16 ***
## t           -0.009849   0.003239  -3.041  0.00246 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 14.05 on 607 degrees of freedom
## Multiple R-squared:  0.015,  Adjusted R-squared:  0.01338
## F-statistic: 9.247 on 1 and 607 DF,  p-value: 0.002461
```

The incertcept for this linear trend:

```r
Hbeta0 = as.numeric(HydroTrend$coefficients[1]) |>
  print()
```

```
## [1] 82.73475
```

The slope for this linear trend:

```r
Hbeta1 = as.numeric(HydroTrend$coefficients[2]) |>
  print()
```

```
## [1] -0.009849298
```

Here, my p-value is less than 0.05 so my coefficient is significant and there is a trend.

Q4

Use the regression coefficients from Q3 to detrend the series. Plot the detrended series and compare with the plots from Q1. What happened? Did anything change?

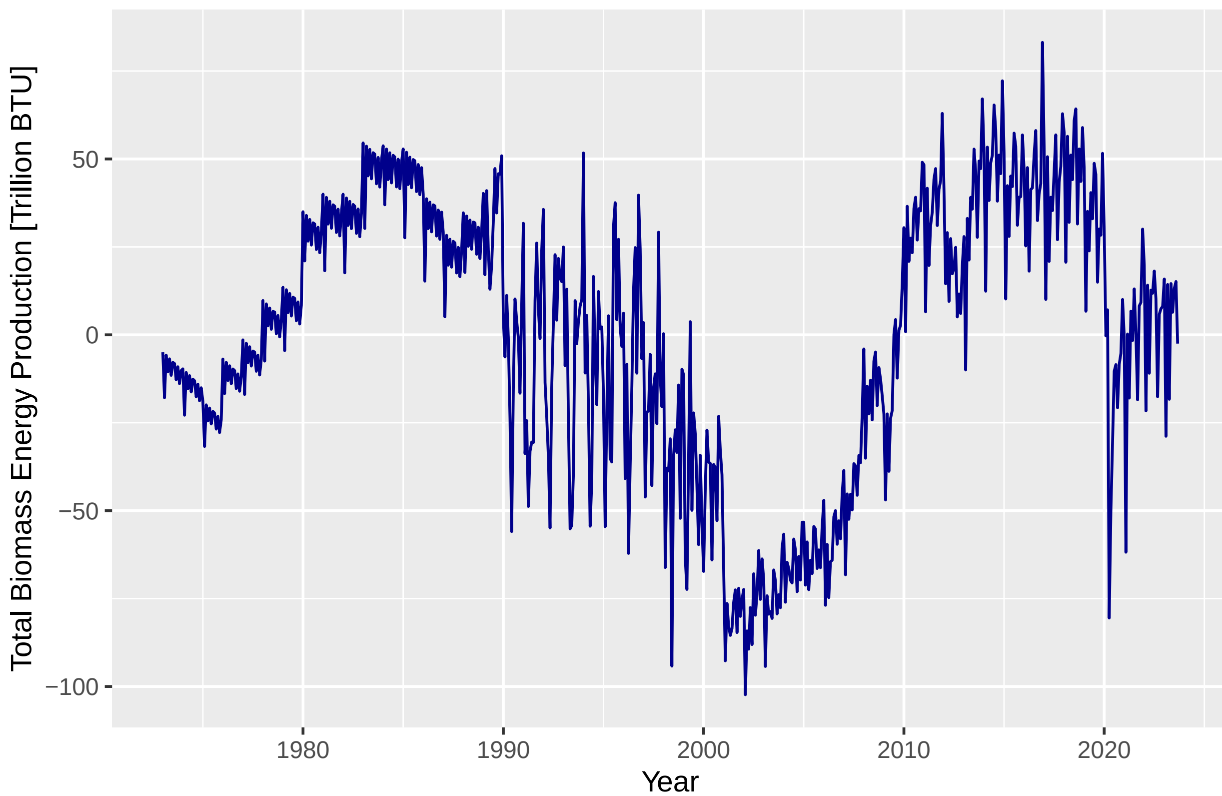
Detrended series = the series - (beta0 + beta1*t)

#Detrended Biomass

```r
DataBiomassTrend <- Bbeta0 + (Bbeta1 * t)
TSDataBiomassTrend <-
  ts(DataBiomassTrend,
     start = c(year1, month1),
     frequency = 12)
```

```r
DataBiomass <- as.numeric(DataBiomass)
DataBiomassTrend <- as.numeric(DataBiomassTrend)
DetrendedBiomass <- DataBiomass - DataBiomassTrend
DetrendedBiomass <-
  ts(DetrendedBiomass,
     start = c(year1, month1),
     frequency = 12)
```

```r
DetrendedBiomass |>
  autoplot(color = "darkblue") +
  ylab("Total Biomass Energy Production [Trillion BTU]") +
  xlab("Year")
```

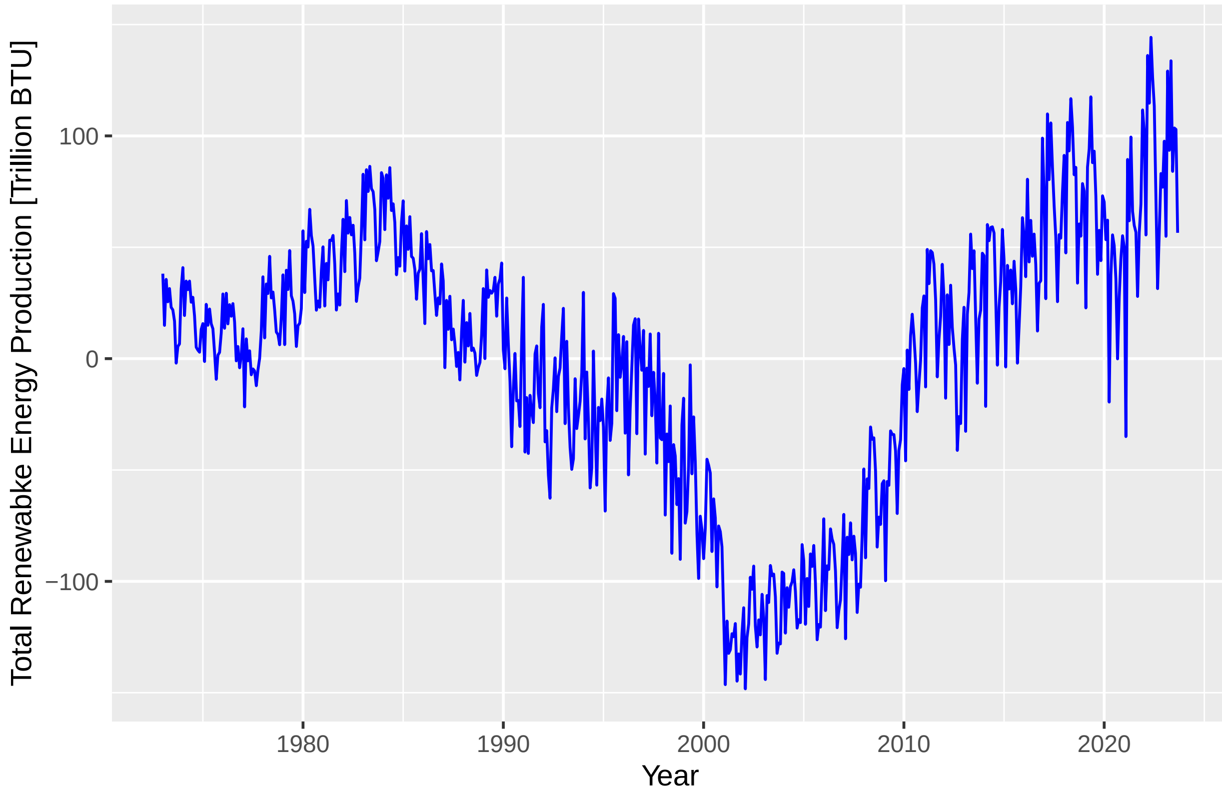


#Detrended Renewable

```
DataRenewableTrend <- Rbeta0 + (Rbeta1 * t)
TSDataRenewableTrend <-
  ts(DataRenewableTrend,
     start = c(year1, month1),
     frequency = 12)
```

```
DataRenewable <- as.numeric(DataRenewable)
DataRenewableTrend <- as.numeric(DataRenewableTrend)
DetrendedRenewable <- DataRenewable - DataRenewableTrend
DetrendedRenewable <-
  ts(DetrendedRenewable,
     start = c(year1, month1),
     frequency = 12)
```

```
DetrendedRenewable |>
  autoplot(color = "blue") +
  ylab("Total Renewabke Energy Production [Trillion BTU]") +
  xlab("Year")
```

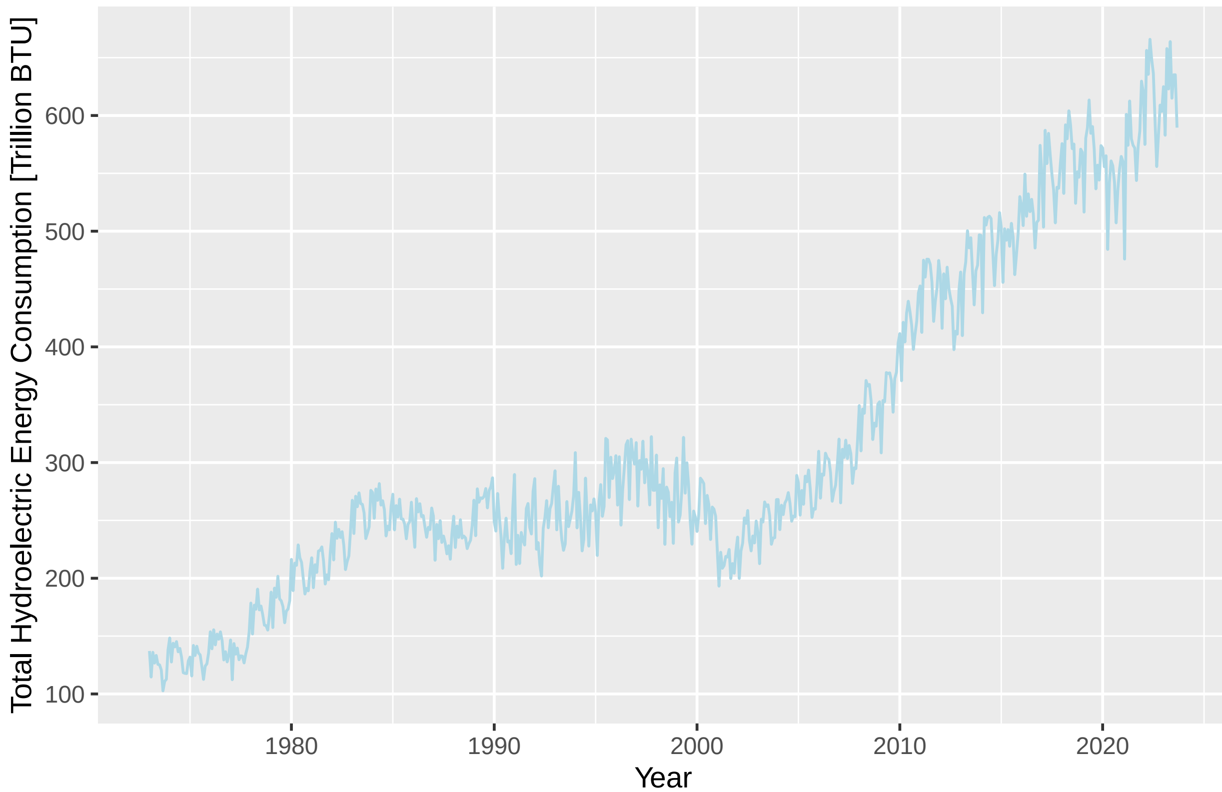


#Detrended Hydroelectric

```
DataHydroTrend <- Hbeta0 + (Hbeta1 * t)
TSDataHydroTrend <-
  ts(DataHydroTrend,
     start = c(year1, month1),
     frequency = 12)
```

```
DataHydro <- as.numeric(DataRenewable)
DataHydroTrend <- as.numeric(DataHydroTrend)
DetrendedHydro <- DataHydro - DataHydroTrend
DetrendedHydro <-
  ts(DetrendedHydro,
     start = c(year1, month1),
     frequency = 12)
```

```
DetrendedHydro |>
  autoplot(color = "lightblue") +
  ylab("Total Hydroelectric Energy Consumption [Trillion BTU]") +
  xlab("Year")
```



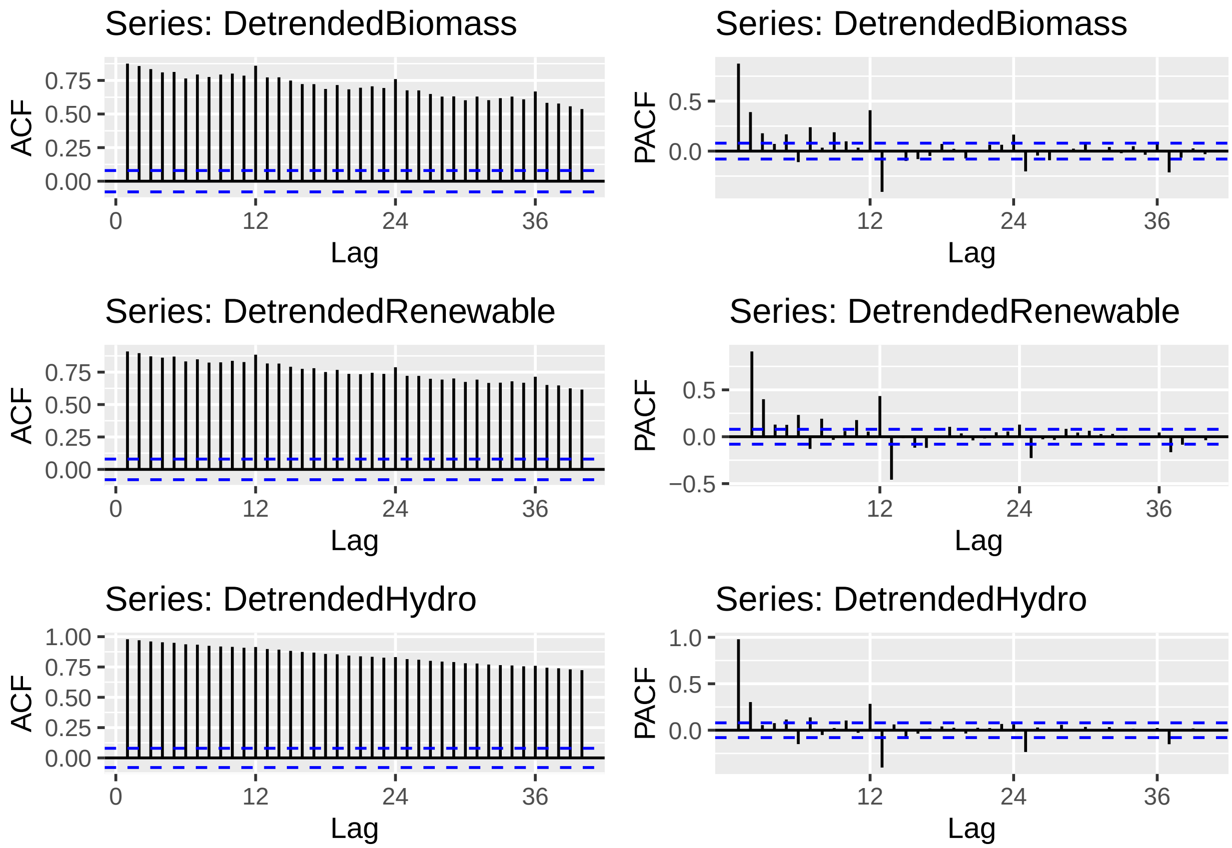
Yes, the plots did change. For biomass and renewable energy, the range is significantly reduced; that is, the ovbious increase in energy consumption as x (time) approaches September of 2023 is largely eliminated and the data is more closely modeled by a horizontal line. The hydroelectric production data is also different; it seems to have an almost linear increase now which is admittedly strange as the data did not have a super ovbious trend be manipulation.

Q5

Plot ACF and PACF for the detrended series and compare with the plots from Q1. You may use plot_grid() again to get them side by side. nut mot mandatory. Did the plots change? How?

```
DBACF <-
  Acf(
    DetrendedBiomass,
    lag.max = 40,
    main = paste("Total Biomass Energy \n Production"),
    plot = FALSE
  )

DBPCF <-
  Pacf(
    DetrendedBiomass,
    lag.max = 40,
    main = paste("Total Biomass Energy \n Production"),
    plot = FALSE
  )
```

```r
DRACF <-
  Acf(
    DetrendedRenewable,
    lag.max = 40,
    main = paste("Total Biomass Energy \n Production"),
    plot = FALSE
  )

DRPCF <-
  Pacf(
    DetrendedRenewable,
    lag.max = 40,
    main = paste("Total Biomass Energy \n Production"),
    plot = FALSE
  )

DHACF <-
  Acf(
    DetrendedHydro,
    lag.max = 40,
    main = paste("Total Biomass Energy \n Production"),
    plot = FALSE
  )

DHPCF <-
  Pacf(
    DetrendedHydro,
    lag.max = 40,
    main = paste("Total Biomass Energy \n Production"),
    plot = FALSE
  )


DB1 <- autoplot(DBACF)
DB2 <- autoplot(DBPCF)
DR1 <- autoplot(DRACF)
DR2 <- autoplot(DRPCF)
DH1 <- autoplot(DHACF)
DH2 <- autoplot(DHPCF)

plot_grid(DB1, DB2, DR1, DR2, DH1, DH2, nrow = 3)
```



Seasonal Component

Set aside the detrended series and consider the original series again from Q1 to answer Q6 to Q8.

Q6

Just by looking at the time series and the acf plots, do the series seem to have a seasonal trend? No need to run any code to answer your question. Just type in you answer below.

Q7

Use function *lm()* to fit a seasonal means model (i.e. using the seasonal dummies) the two time series. Ask R to print the summary of the regression. Interpret the regression output. From the results which series have a seasonal trend? Do the results match you answer to Q6?

Q8

Use the regression coefficients from Q7 to deseason the series. Plot the deseason series and compare with the plots from part Q1. Did anything change?

Q9

Plot ACF and PACF for the deseason series and compare with the plots from Q1. You may use plot_grid() again to get them side by side. nut mot mandatory. Did the plots change? How?