ENV 790.30 - Time Series Analysis for Energy Data | Spring 2022

Assignment 3 - Due date 02/08/22

Narissa Jimenez-Petchumrus

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 project open the first thing you will do is change "Student Name" on line 3 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. Rename the pdf file such that it includes your first and last name (e.g., "LuanaLima_TSA_A03_Sp22.Rmd"). Submit this pdf using Sakai.

Questions

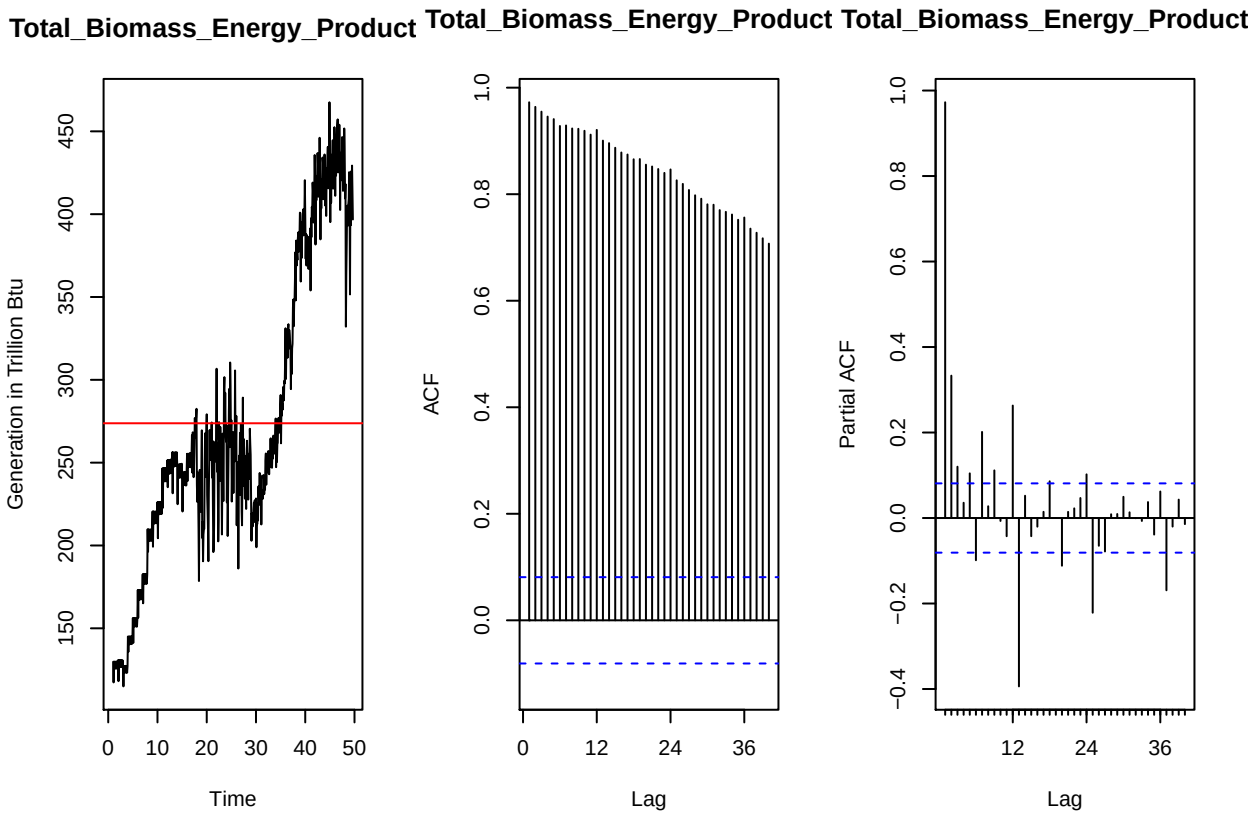
Consider the same data you used for A2 from the spreadsheet "Table_10.1_Renewable_Energy_Production_and_Consumption
The data comes from the US Energy Information and Administration and corresponds to the January 2022 **Monthly** Energy Review. Once again you will work only with the following columns: Total Biomass Energy Production, Total Renewable Energy Production, Hydroelectric Power Consumption. Create a data frame structure with these three time series only.

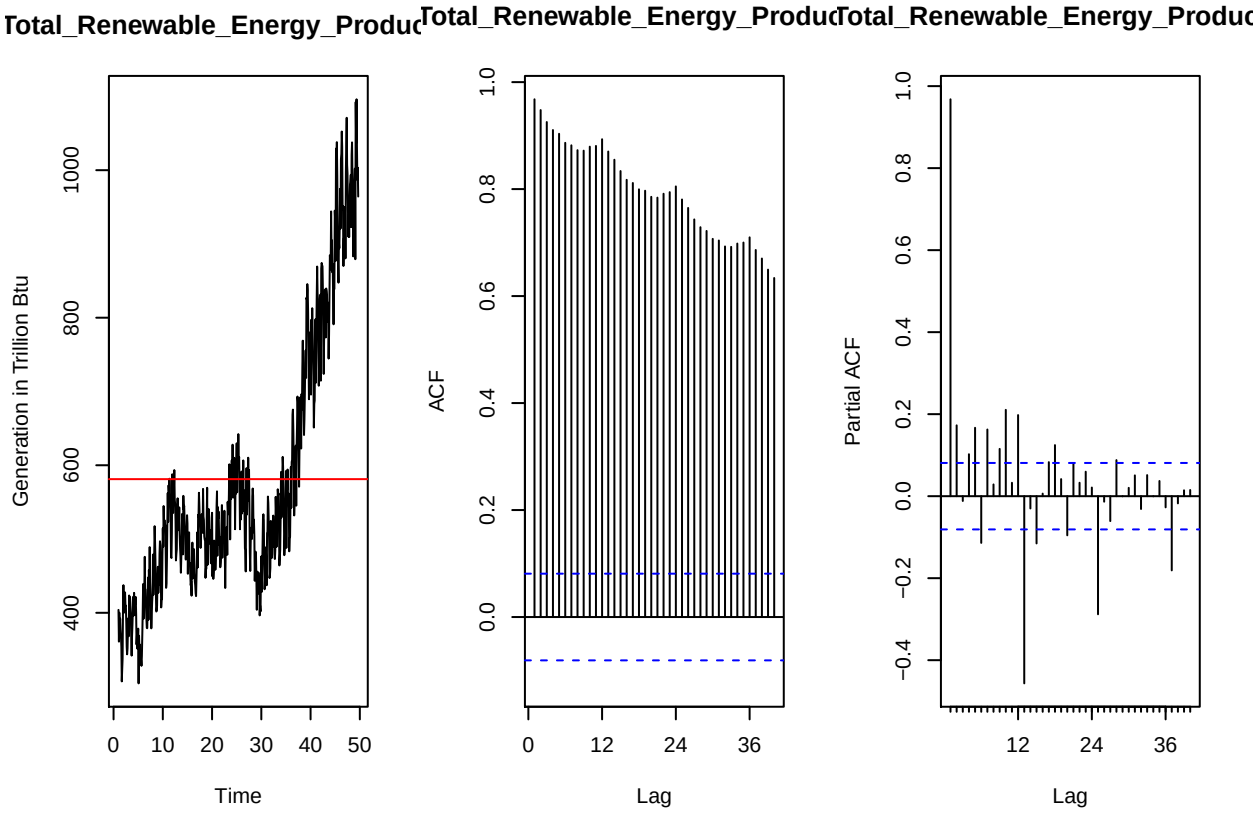
R packages needed for this assignment:"forecast","tseries", and "Kendall". Install these packages, if you haven't done yet. Do not forget to load them before running your script, since they are NOT default packages.\

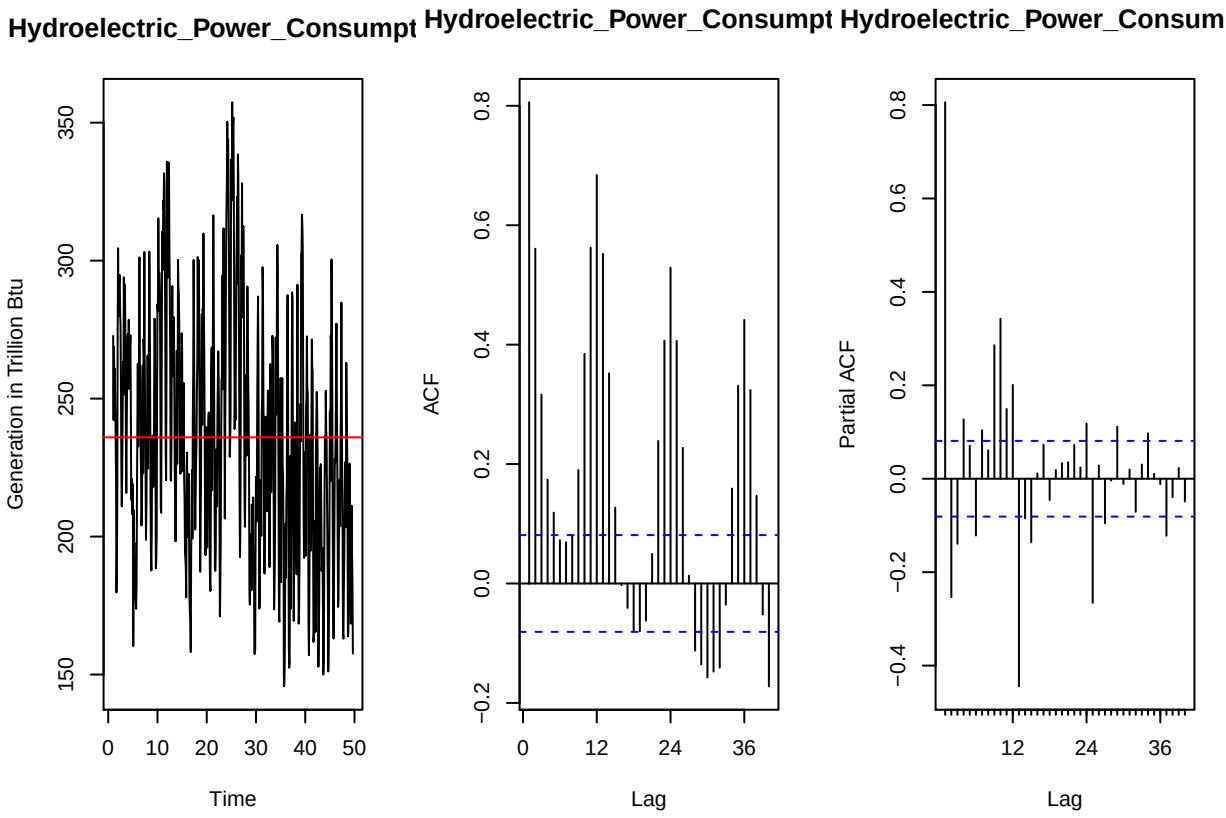
##Trend Component

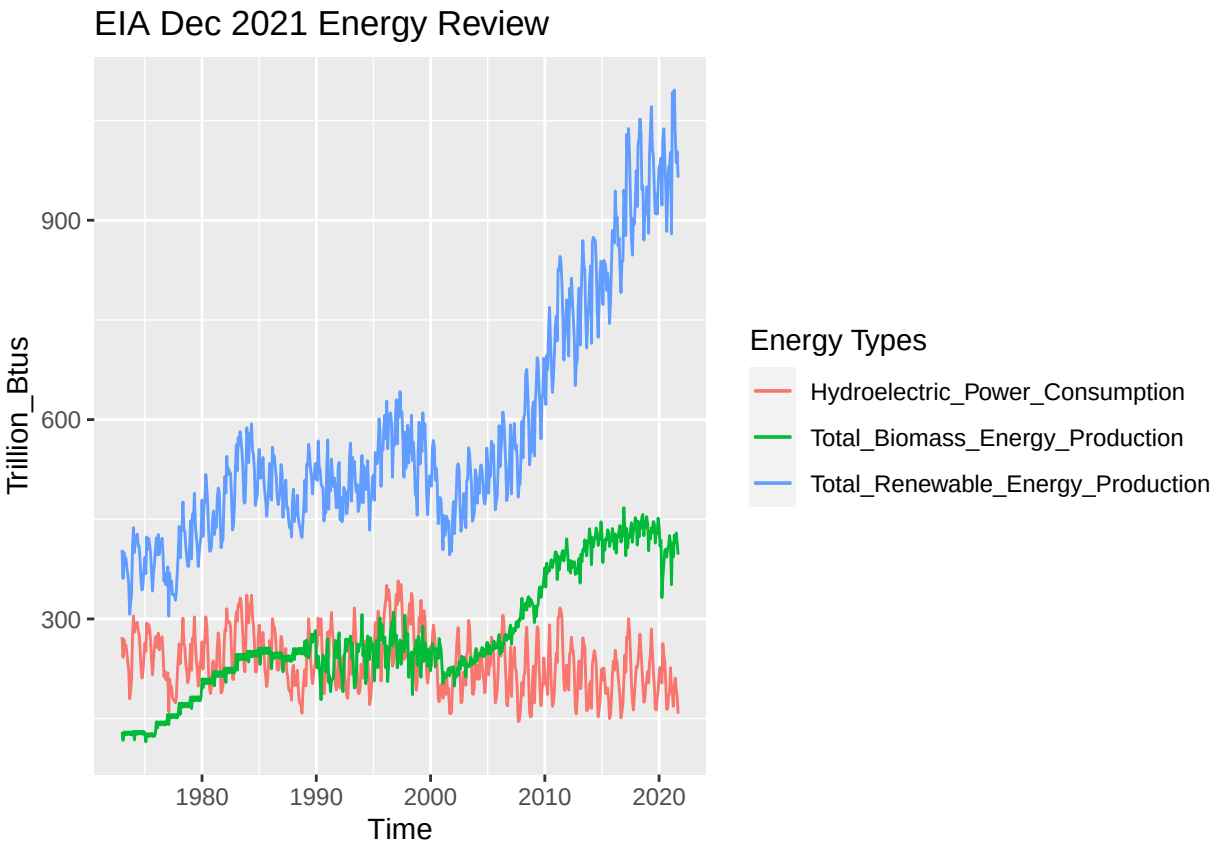
Q1

Create a plot window that has one row and three columns. And then for each object on your data frame, fill the plot window with time series plot, ACF and PACF. You may use the some code form A2, but I want all three plots on the same window this time. (Hint: use par() function)


Total_Biomass_Energy_Product
Generation in Trillion Btu
Time
Total_Biomass_Energy_Product
ACF
Lag
Total_Biomass_Energy_Product
Partial ACF
Lag


Total_Renewable_Energy_Produc
Generation in Trillion Btu
Time
Total_Renewable_Energy_Produc
ACF
Lag
Total_Renewable_Energy_Produc
Partial ACF
Lag


Hydroelectric_Power_Consumpt
Generation in Trillion Btu
Time
Hydroelectric_Power_Consumpt
ACF
Lag
Hydroelectric_Power_Consumpt
Partial ACF
Lag


EIA Dec 2021 Energy Review
Trillion_Btus
Time
Energy Types
Hydroelectric_Power_Consumption
Total_Biomass_Energy_Production
Total_Renewable_Energy_Production

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?

From the plots in Q1, it appears that both biomass and renewable energy production both are upward trending over time. Hydroelectric power consumption doesn't appear to have a trend.

Q3

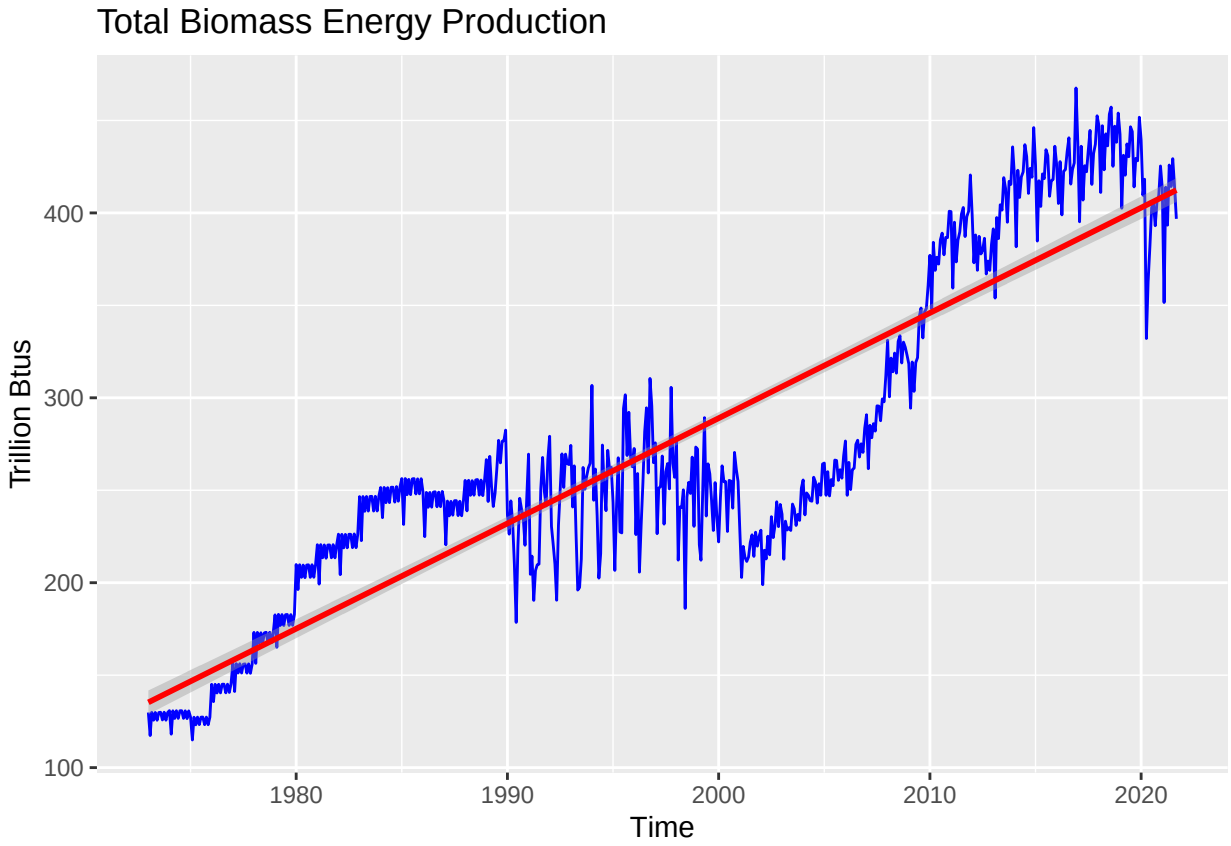
Use the *lm()* function to fit a linear trend to the three time series. Ask R to print the summary of the regression. Interpret the regression output, i.e., slope and intercept. Save the regression coefficients for further analysis.

For Biomass, the intercept = 134.78 and the slope = 0.474. This shows that approximately for every month, there is an increase of 0.474 trillion Btus in Biomass energy generation.

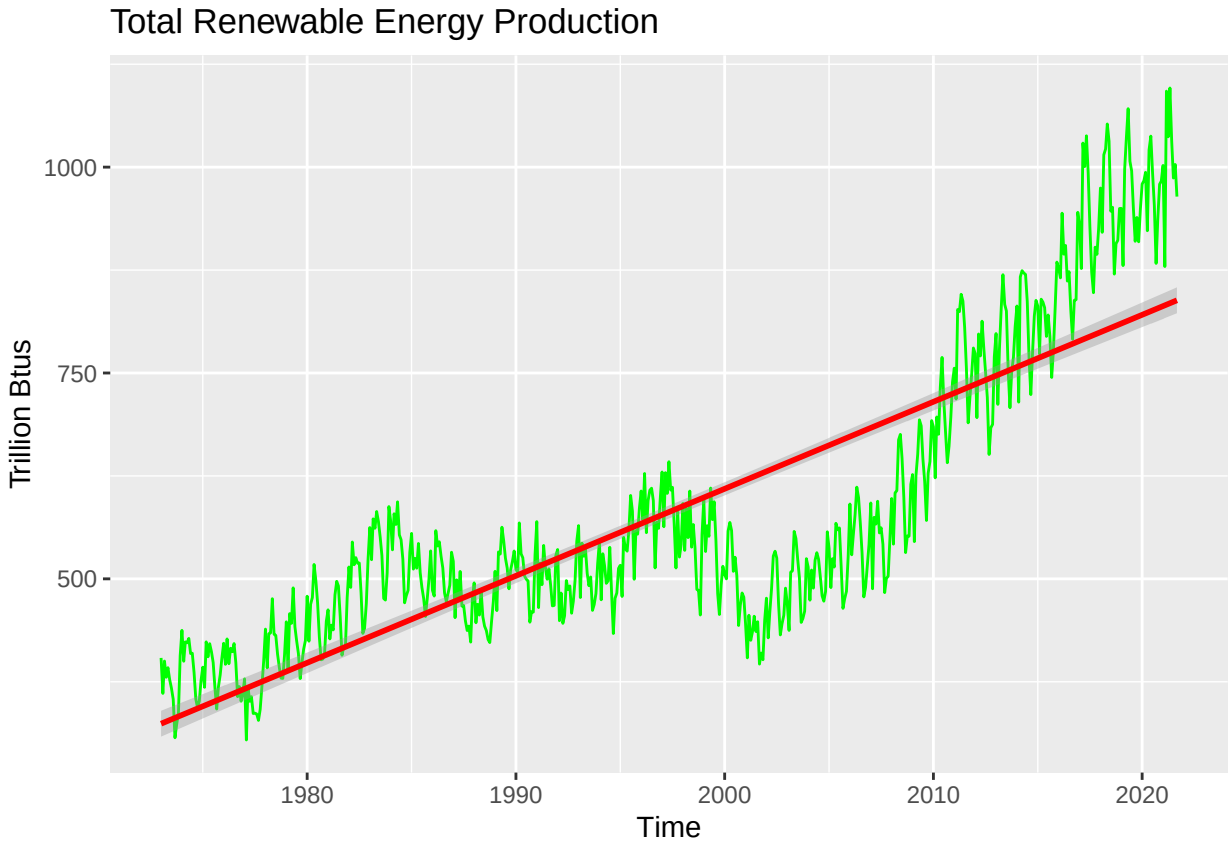
For Hydroelectric, the intercept = 259.18 and the slope = -0.079. This shows that approximately for every month, there is a decrease of 0.079 trillion Btus in Hydro energy consumption.

For Renewables, the intercept = 323.18 and the slope = 0.88. This shows that approximately for every month, there is a decrease of 0.88 trillion Btus in Renewable energy generation.

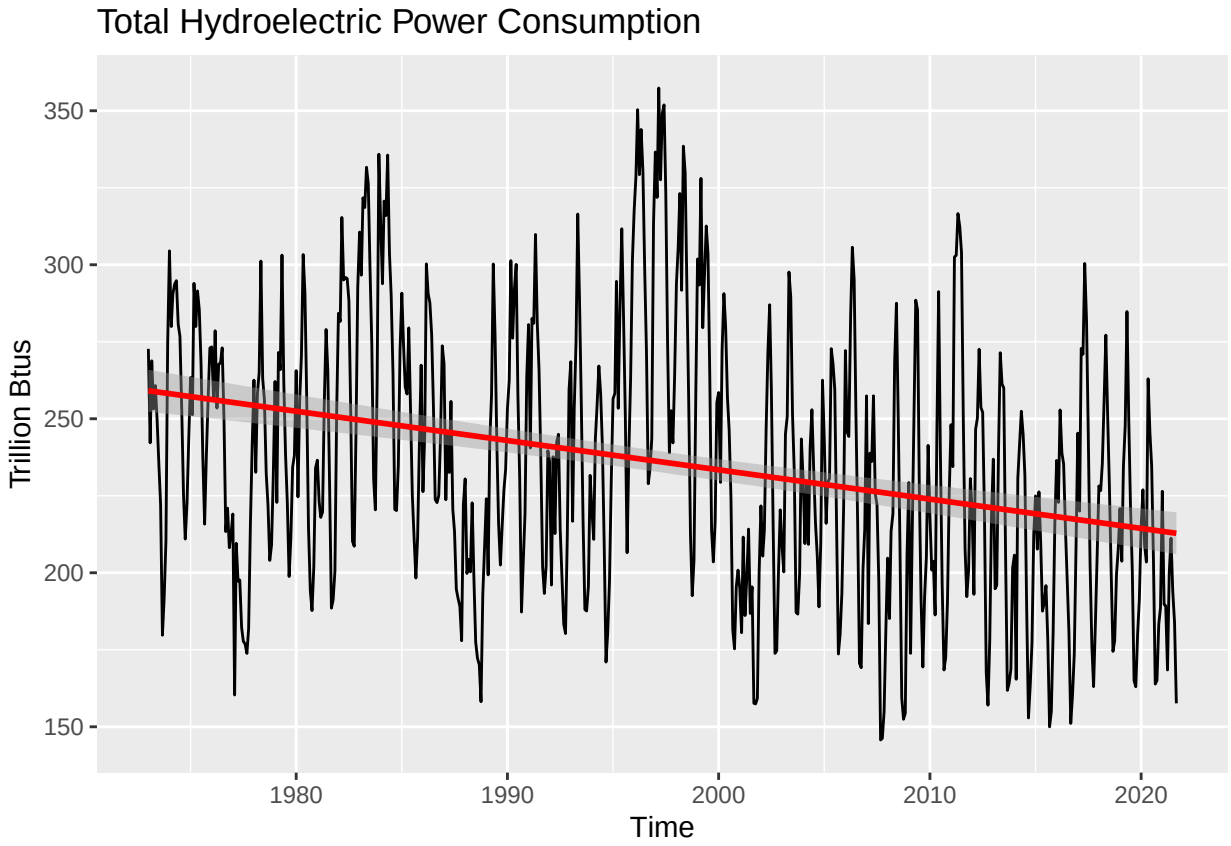
```
##
## Call:
## lm(formula = Total_Biomass_Energy_Production ~ t, data = energy_data2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -101.892  -24.306    4.932   33.103   82.292
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.348e+02  3.282e+00   41.07   <2e-16 ***
## t           4.744e-01  9.705e-03   48.88   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 39.64 on 583 degrees of freedom
## Multiple R-squared:  0.8039, Adjusted R-squared:  0.8035
## F-statistic:  2389 on 1 and 583 DF,  p-value: < 2.2e-16
```




```
##
## Call:
## lm(formula = Total_Renewable_Energy_Production ~ t, data = energy_data2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -230.488  -57.869    5.595   62.090  261.349
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 323.18243    8.02555   40.27   <2e-16 ***
## t             0.88051    0.02373   37.10   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 96.93 on 583 degrees of freedom
## Multiple R-squared:  0.7025, Adjusted R-squared:  0.702
## F-statistic:  1377 on 1 and 583 DF,  p-value: < 2.2e-16
```




```
##
## Call:
## lm(formula = Hydroelectric_Power_Consumption ~ t, data = energy_data2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -94.892 -31.300  -2.414  27.876 121.263
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 259.18303    3.47464  74.593  < 2e-16 ***
## t            -0.07924    0.01027  -7.712 5.36e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 41.97 on 583 degrees of freedom
## Multiple R-squared:  0.09258,    Adjusted R-squared:  0.09103
## F-statistic: 59.48 on 1 and 583 DF,  p-value: 5.364e-14
```



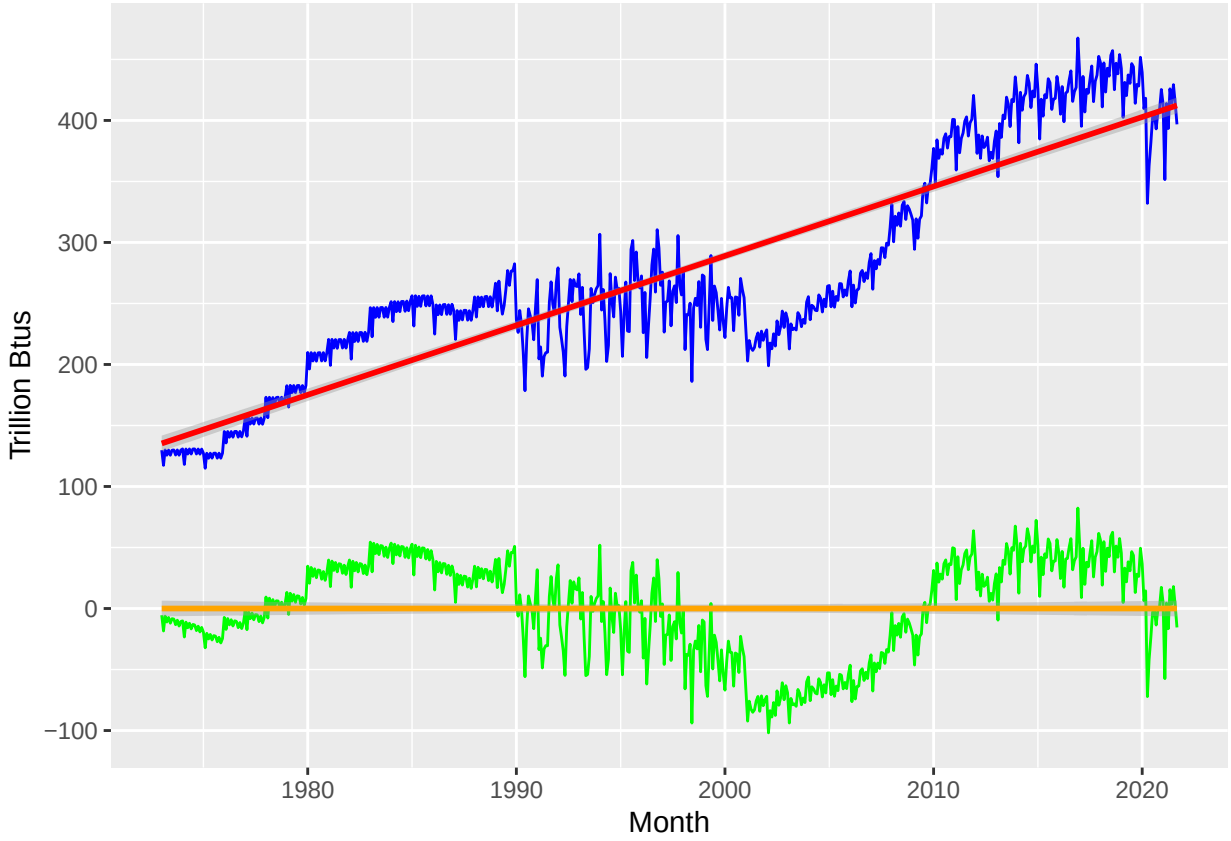
Q4

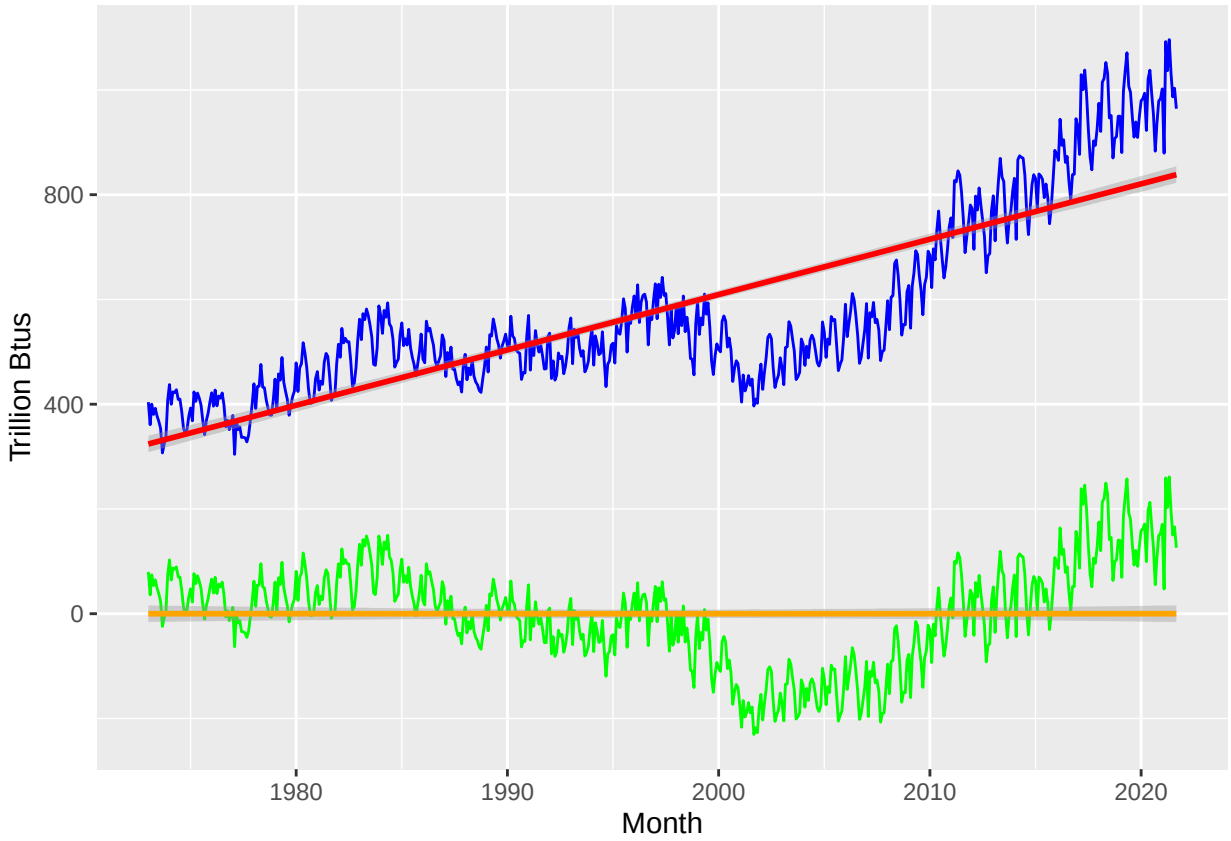
Use the regression coefficients from Q3 to detrend the series. Plot the detrended series and compare with the plots from Q1. What happened? Did anything change?

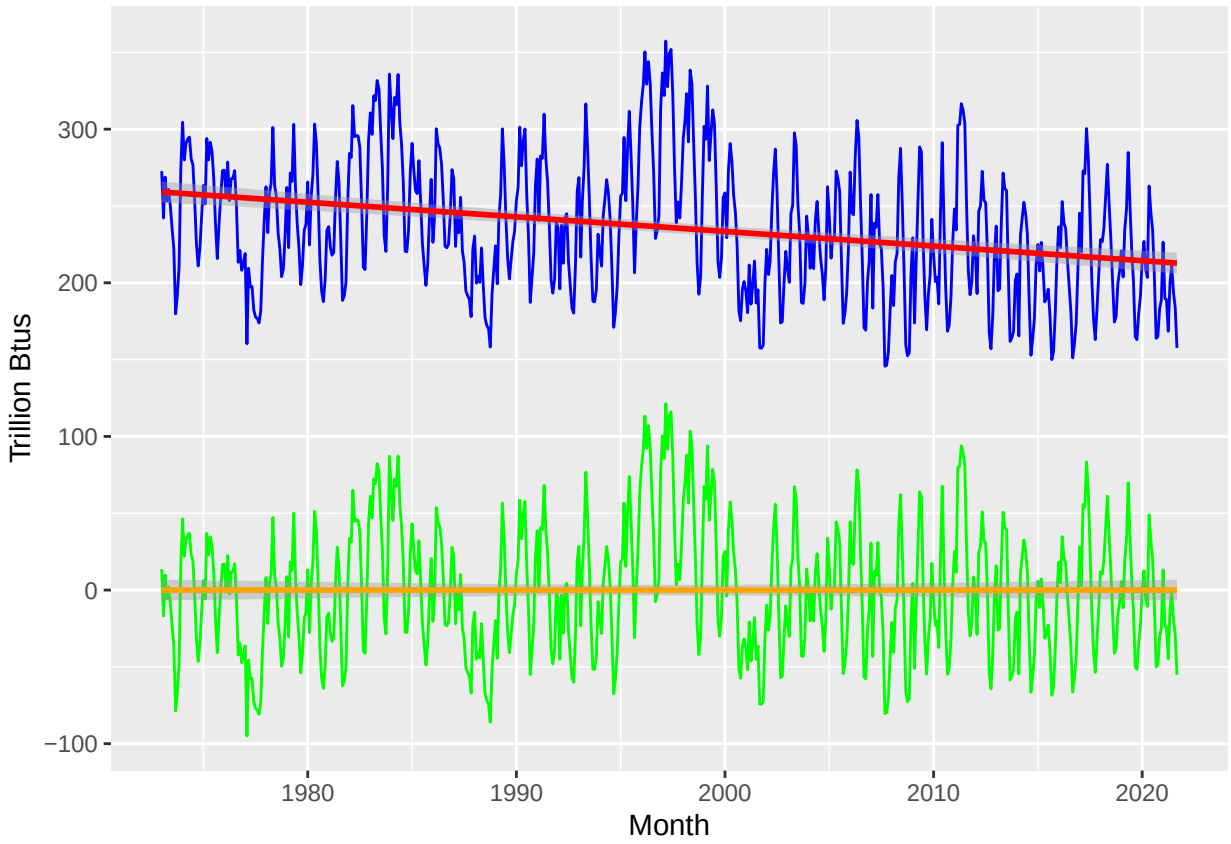
Regarding biomass, the detrended series isn't showing that dramatic of an upward trend over time. In addition the mean of the detrended series is approximately at y=0. It also appears easier to see seasonality if there is any.

Regarding renewables, the detrended series is also not showing that dramatic of an upward trend over time. The mean of this detrended series is also at approximately y=0. It also appears easier to see seasonality if there is any.

Regarding hydroelectric, the detrended series went from a slightly negative trend (based on just staring at the graph though) to showing no potential trend (thus detrended but again this is by only looking at the graph). The mean of this detrended series is also approximately at y=0. It also appears easier to see seasonality if there is any.


400
300
200
100
0
−100
Trillion Btus
1980
1990
2000
2010
2020
Month

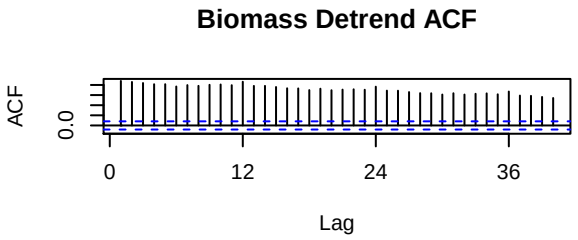

800
400
0
Trillion Btus
1980
1990
2000
2010
2020
Month

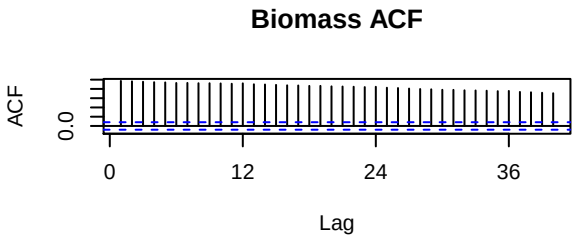


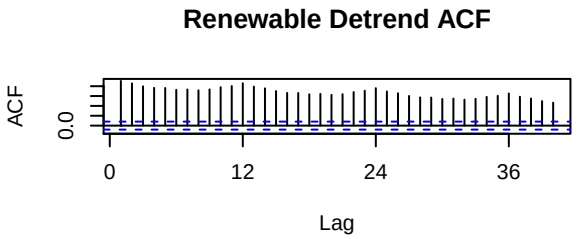
Q5

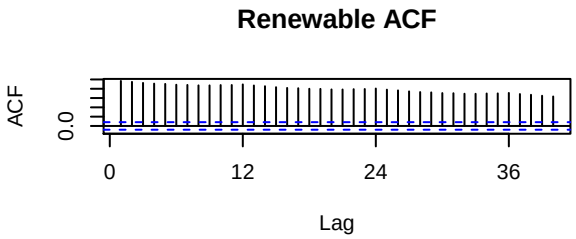
Plot ACF and PACF for the detrended series and compare with the plots from Q1. Did the plots change? How?

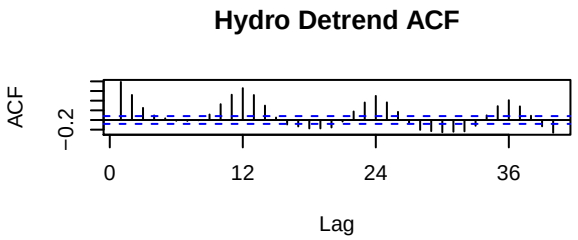
Looking at the plots comparing PACF and ACF of the trend and detrended series, there isn't that much of a dramatic change between the ACFs and PACFs. The only ones that appear to be different between the trend and detrended series are renewables ACF (the detrended renewables ACF shows a little seasonality compared to trend renewable ACF), biomass ACF (the detrended ACF shos a little less of a gradual decline compared to the trend ACF, potentially some seasonality), biomass PACF (lag 12 on the detrended series looks barely a little higher unless my eyesight sucks), and renewables PACF (a few of the lags on the detrended PACF looks taller compared to the trend PACF which again unless my eyesight sucks).

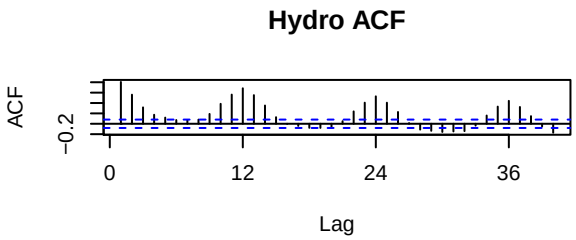

Biomass Detrend ACF
ACF
0.0
0
12
24
36
Lag

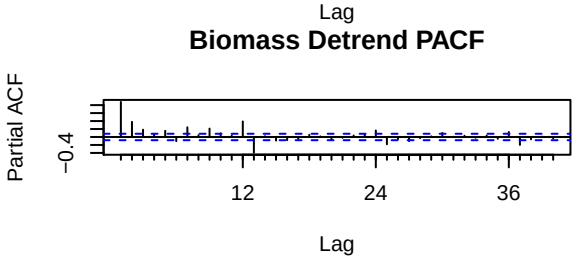

Biomass ACF
ACF
0.0
0
12
24
36
Lag

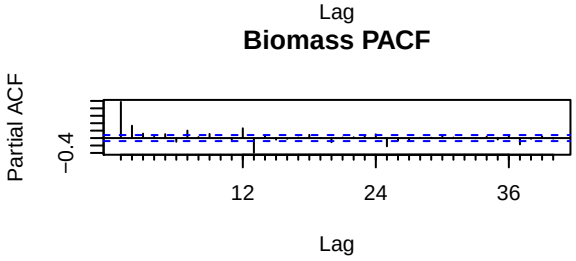

Renewable Detrend ACF
ACF
0.0
0
12
24
36
Lag

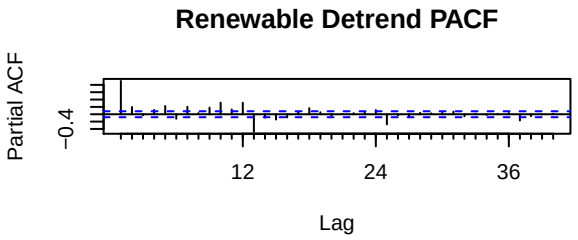

Renewable ACF
ACF
0.0
0
12
24
36
Lag

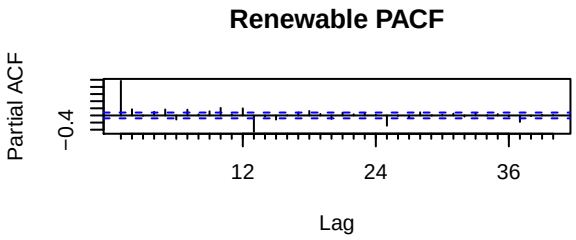

Hydro Detrend ACF
ACF
−0.2
0
12
24
36
Lag

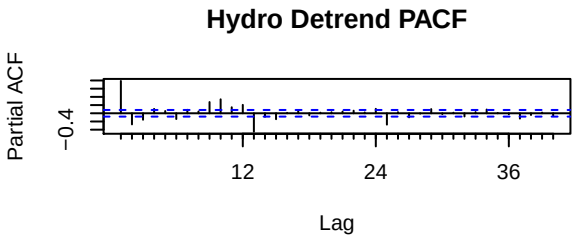

Hydro ACF
ACF
−0.2
0
12
24
36
Lag

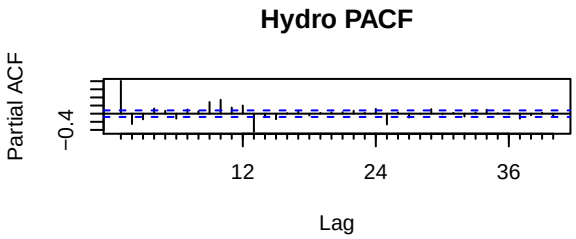

Biomass Detrend PACF
Partial ACF
−0.4
12
24
36
Lag


Biomass PACF
Partial ACF
−0.4
12
24
36
Lag


Renewable Detrend PACF
Partial ACF
−0.4
12
24
36
Lag


Renewable PACF
Partial ACF
−0.4
12
24
36
Lag


Hydro Detrend PACF
Partial ACF
−0.4
12
24
36
Lag


Hydro PACF
Partial ACF
−0.4
12
24
36
Lag

Seasonal Component

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

Q6

Do the series seem to have a seasonal trend? Which serie/series? Use function *lm()* to fit a seasonal means model (i.e. using the seasonal dummies) to this/these time series. Ask R to print the summary of the regression. Interpret the regression output. Save the regression coefficients for further analysis.

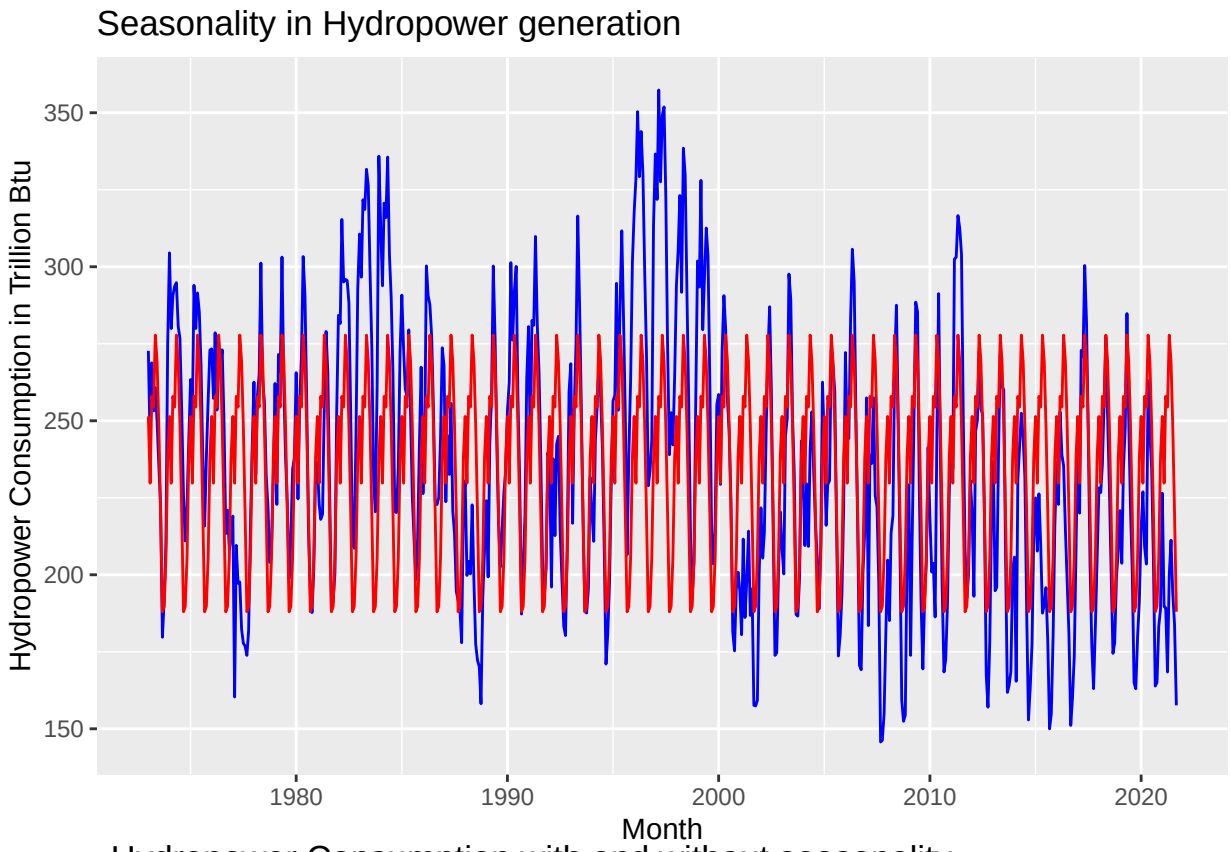
The series that has an obvious seasonal trend is hydroelectric based on plots and graphs. During mostly non-summer months, there appears to be more hydroelectric resources available as the regression coefficients are higher as there's more precipitation around this time. Late summer to fall there appears to be less hydroelectric sources available given these months are dryer.

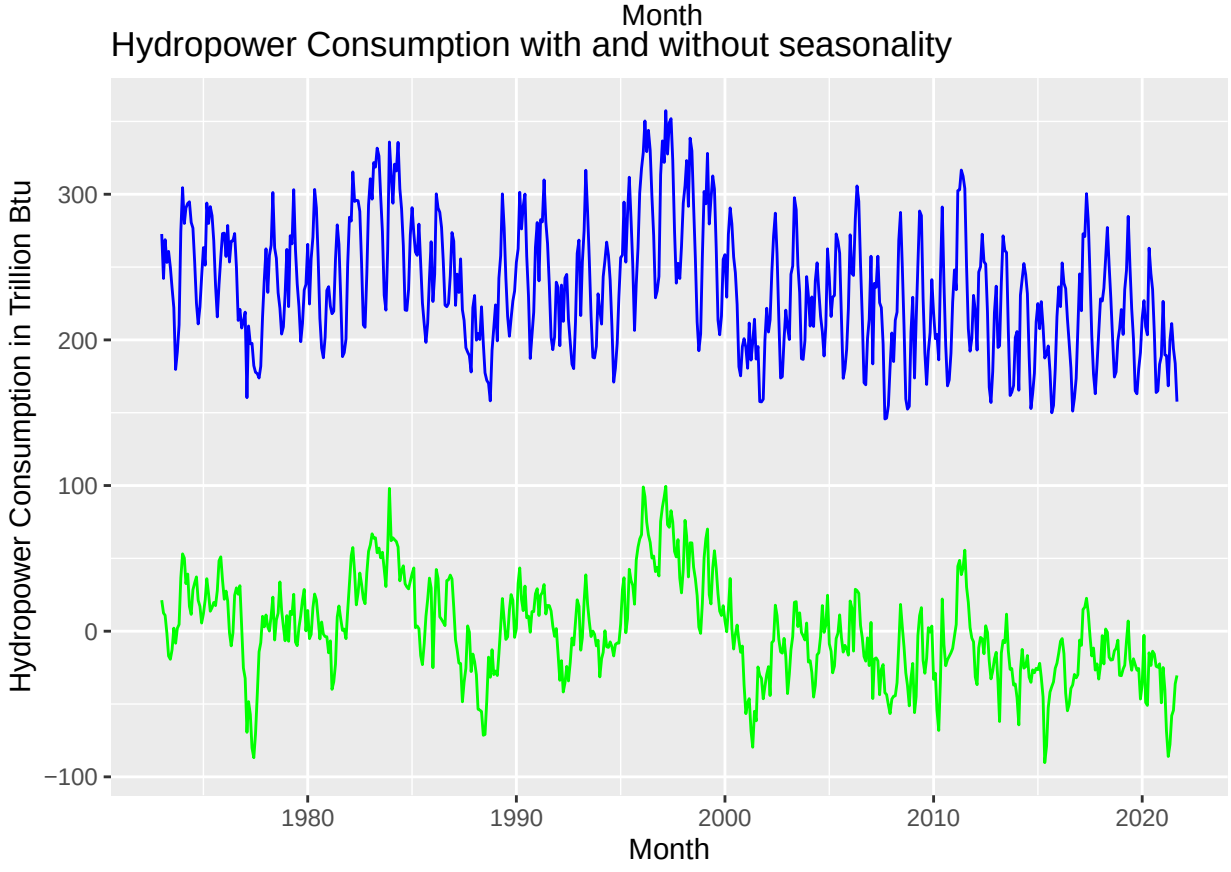
```
##
## Call:
## lm(formula = energy_data2[, 4] ~ dummies)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -90.253 -23.017  -3.042  21.487  99.478
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  237.841      4.892  48.616  < 2e-16 ***
## dummiesJan    13.558      6.883   1.970  0.04936 *
## dummiesFeb    -8.090      6.883  -1.175  0.24037
## dummiesMar    20.067      6.883   2.915  0.00369 **
## dummiesApr    16.619      6.883   2.414  0.01607 *
## dummiesMay    39.961      6.883   5.805 1.06e-08 ***
## dummiesJun    31.315      6.883   4.549 6.57e-06 ***
## dummiesJul    10.511      6.883   1.527  0.12732
## dummiesAug   -17.853      6.883  -2.594  0.00974 **
## dummiesSep   -49.852      6.883  -7.242 1.43e-12 ***
## dummiesOct   -48.086      6.919  -6.950 9.96e-12 ***
## dummiesNov   -32.187      6.919  -4.652 4.08e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 33.89 on 573 degrees of freedom
## Multiple R-squared:  0.4182, Adjusted R-squared:  0.4071
## F-statistic: 37.45 on 11 and 573 DF,  p-value: < 2.2e-16
```

Q7

Use the regression coefficients from Q6 to deseason the series. Plot the deseason series and compare with the plots from part Q1. Did anything change?

Comparing the seasonal and deseasoned data of hydroelectric power, the deseasoned data shows some values that could potentially be outliers that the seasonal data doesn't reveal. In addition, the mean appears to be at y=0.


Seasonality in Hydropower generation
Hydropower Consumption in Trillion Btu
350
300
250
200
150
1980
1990
2000
2010
2020
Month


Hydropower Consumption with and without seasonality
Hydropower Consumption in Trillion Btu
300
200
100
0
−100
1980
1990
2000
2010
2020
Month

Q8

Plot ACF and PACF for the deseason series and compare with the plots from Q1. Did the plots change? How?

Seasonality for hydroelectric appears to be much more controlled for comparing the deaseasoned ACF with the seasonal ACF. Regarding the PACFs, the deseasoned hydroelectric graph shows shorter lags compared to the seasonal PACF.