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

Assignment 3 - Due date 02/10/23

Zhengqi Jiao

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_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).

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(Kendall)
library(lubridate)
```

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

## The following objects are masked from 'package:base':
##
```

```
##     date, intersect, setdiff, union
```

```
library(ggplot2)
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(tidyr)
```

##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)

```
#Importing data set
energy_data <- read.csv(file="/Users/christine/Documents/TimeSeriesAnalysis/TimeSeriesAnalysis_Jiao/Data
```

```
energy_data1 <- energy_data[,4:6]
head(energy_data1)
```

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

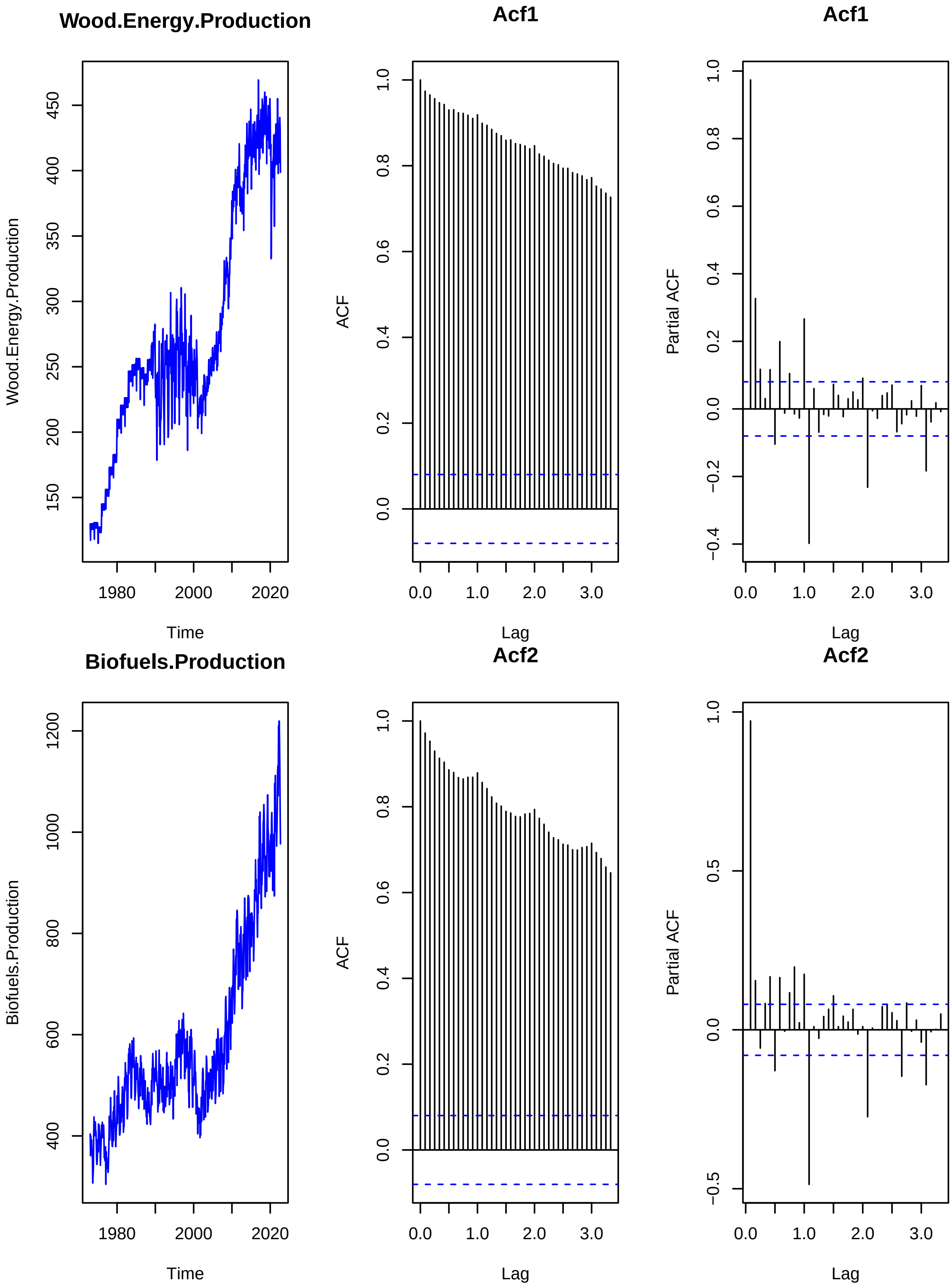
```
ts_energy_data1 <- ts(energy_data1,frequency=12,start=c(1973,1))
head(ts_energy_data1,20)
```

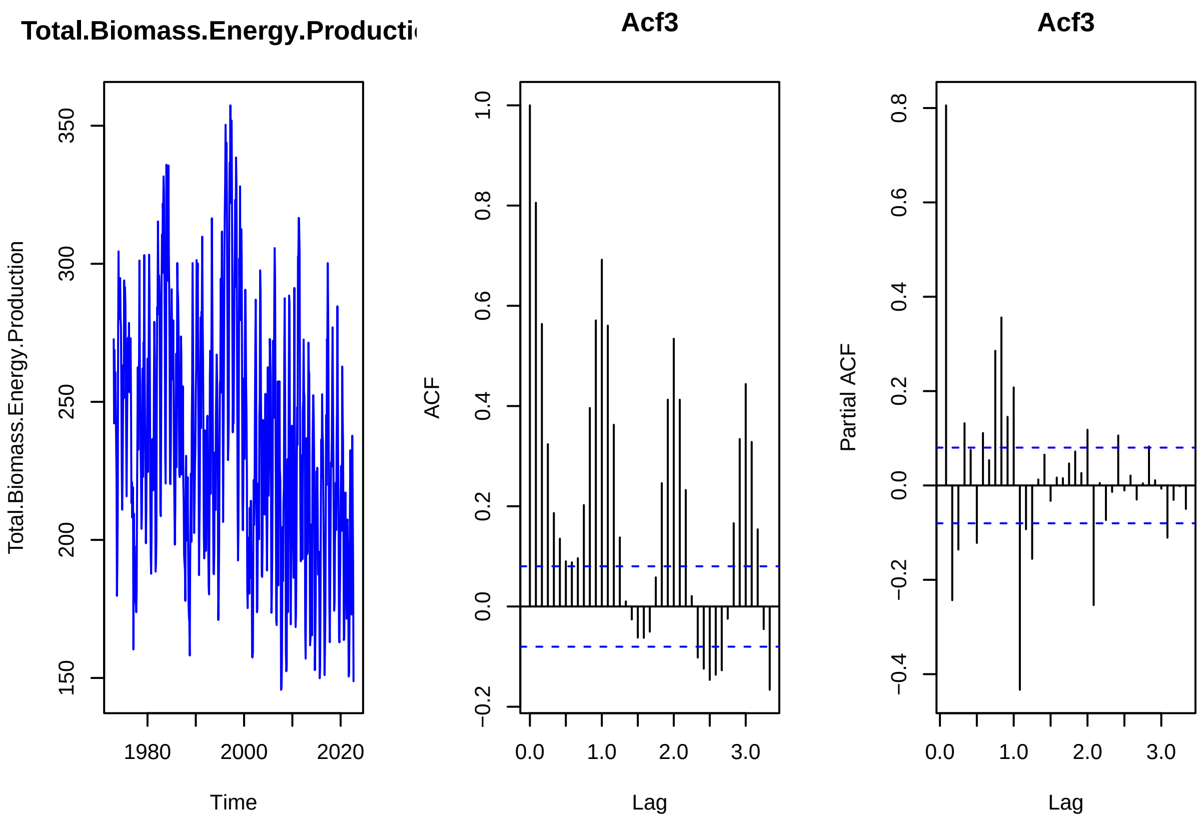
```
##          Total.Biomass.Energy.Production Total.Renewable.Energy.Production
## Jan 1973                         129.787                           403.981
## Feb 1973                         117.338                           360.900
## Mar 1973                         129.938                           400.161
## Apr 1973                         125.636                           380.470
## May 1973                         129.834                           392.141
## Jun 1973                         125.611                           377.232
## Jul 1973                         129.787                           367.325
```

```
## Aug 1973                       129.918                               353.757
## Sep 1973                       125.782                               307.006
## Oct 1973                       129.970                               323.453
## Nov 1973                       125.643                               337.817
## Dec 1973                       129.824                               406.694
## Jan 1974                       130.807                               437.467
## Feb 1974                       118.091                               399.942
## Mar 1974                       130.727                               423.474
## Apr 1974                       126.583                               422.323
## May 1974                       130.789                               427.657
## Jun 1974                       126.611                               409.281
## Jul 1974                       130.756                               409.719
## Aug 1974                       130.763                               386.101
##          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
## Jul 1973                       235.670
## Aug 1973                       222.077
## Sep 1973                       179.733
## Oct 1973                       191.723
## Nov 1973                       210.285
## Dec 1973                       274.435
## Jan 1974                       304.506
## Feb 1974                       279.950
## Mar 1974                       290.582
## Apr 1974                       293.702
## May 1974                       294.828
## Jun 1974                       280.695
## Jul 1974                       276.772
## Aug 1974                       253.175
```

```
my_date11 <- paste(energy_data[,1],ts_energy_data1[,2],sep="-")

ed <- energy_data %>%
   separate(Month, c('year', 'month'))

my_date <- paste(ed[,2],ed[,1],sep="-")
my_date <- my(my_date)
```

```
names <- colnames(energy_data)
for(i in 1:3){
   par(mfrow=c(1,3))
   plot(ts_energy_data1[,i],type="l",col="blue",ylab=paste0(names[i+1]),main=names[i+1])
   acf(ts_energy_data1[,i],lag.max = 40, main=paste0("Acf",i))
   pacf(ts_energy_data1[,i],lag.max = 40, main=paste0("Acf",i))
   #Acf(ts_energy_data1[,i],lag.max = 40,main = paste0 ("ACF of ",names[i+1]), ylim=c(-1,1))
   #Pacf(ts_energy_data1[,i],lag.max = 40,main=paste0("PACF of",names[i+1]))
   }
```


Wood.Energy.Production
Wood.Energy.Production
Time
Acf1
ACF
Lag
Acf1
Partial ACF
Lag
Biofuels.Production
Biofuels.Production
Time
Acf2
ACF
Lag
Acf2
Partial ACF
Lag




```
#acf(ts_energy_data1[,1],lag.max = 40, main = "Total.Biomass.Energy.Production")
#pacf(ts_energy_data1[,1],lag.max = 40)
#Acf(ts_inflow_data[,i],lag.max = 40,main=names[i+1])
   #pacf(ts_inflow_data[,i],lag.max = 40,main=paste0("HP",i))
```

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?

The Total Biomass Energy Production has no clear uptrend and downtrend. So it is the stationary trend. Both of the Total Renewable Energy Production and the Hydroelectric Power Consumption apear to have a upward pattern in time series. So the Total Renewable Energy Production and the Hydroelectric Power Consumption is upward trend or uptrend.

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.

```
nobs <- nrow(ts_energy_data1)
t <- 1:nobs
iHP <- 0
nobs
```

```
## [1] 597
```

```
#Fit a linear trend to TS of iHP
Linear_trend0 <- lm(ts_energy_data1[,iHP+1] ~ t)
```

```r
summary(Linear_trend0)
```

```
##
## Call:
## lm(formula = ts_energy_data1[, iHP + 1] ~ t)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -102.800  -23.994    5.667   32.265   82.192
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.337e+02  3.245e+00   41.22   <2e-16 ***
## t           4.800e-01  9.402e-03   51.05   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 39.59 on 595 degrees of freedom
## Multiple R-squared:  0.8142, Adjusted R-squared:  0.8138
## F-statistic:  2607 on 1 and 595 DF,  p-value: < 2.2e-16
```

The intercept is 1.337e+02, and the slope is 4.800e-01. The relationship here is positive. The p-value here is less than 2.2e-16, which also smaller than 0.05. Therefore, the null hypothesis is rejected, and there is a trend in Biomass energy production.

```r
nobs <- nrow(ts_energy_data1)
t <- 1:nobs
iHP <- 1

#Fit a linear trend to TS of iHP
Linear_trend1 <- lm(ts_energy_data1[,iHP+1] ~ t)
summary(Linear_trend1)
```

```
##
## Call:
## lm(formula = ts_energy_data1[, iHP + 1] ~ t)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -238.75  -61.85    8.59   64.48  352.27
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 312.2475     8.4902   36.78   <2e-16 ***
## t             0.9362     0.0246   38.05   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 103.6 on 595 degrees of freedom
## Multiple R-squared:  0.7088, Adjusted R-squared:  0.7083
## F-statistic:  1448 on 1 and 595 DF,  p-value: < 2.2e-16
```

The intercept is 312.2475, and the slope is 0.9362. The relationship here is positive. The p-value here is less than 2.2e-16, which also smaller than 0.05, therefore, the null hypothesis is rejected, and there is a trend in renewable energy production.

```r
nobs <- nrow(ts_energy_data1)
t <- 1:nobs
iHP <- 2

#Fit a linear trend to TS of iHP
Linear_trend2 <- lm(ts_energy_data1[,iHP+1] ~ t)
summary(Linear_trend2)
```

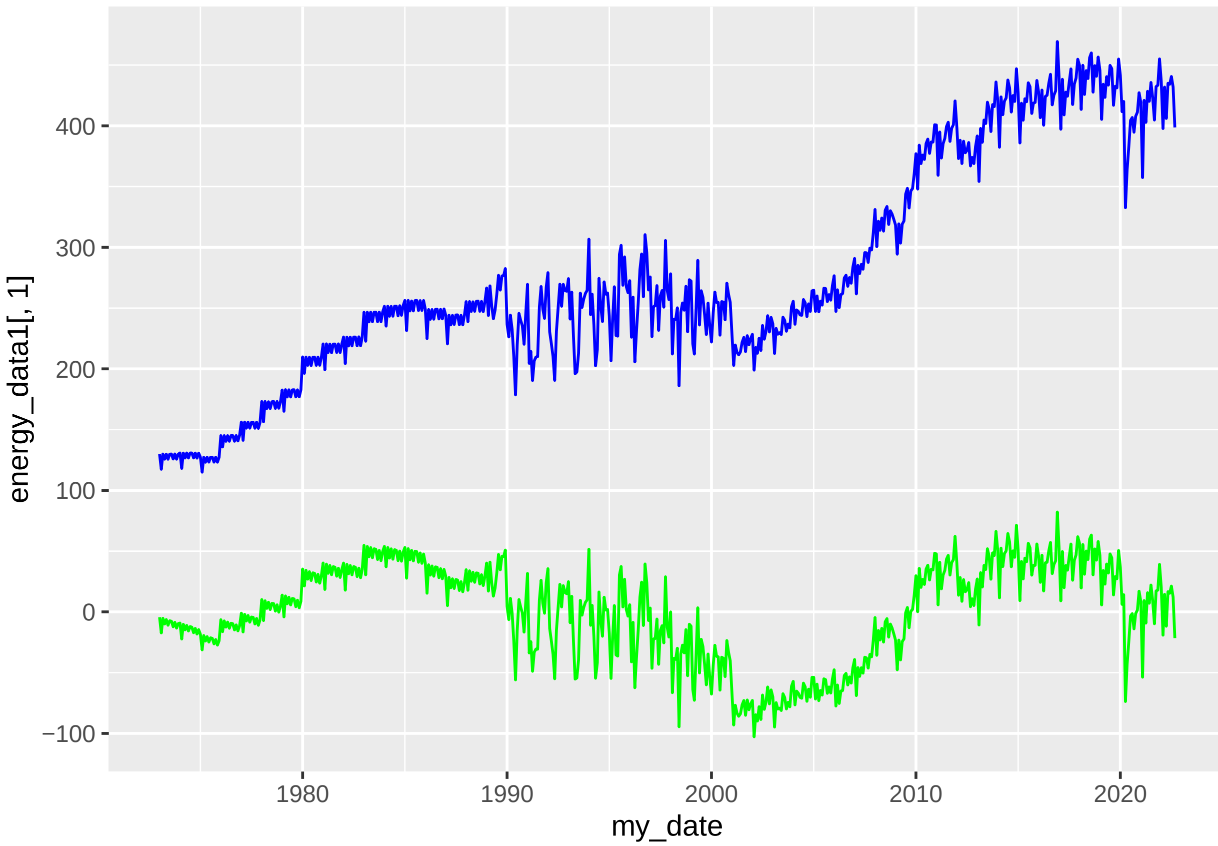
```
##
## Call:
## lm(formula = ts_energy_data1[, iHP + 1] ~ t)
##
## Residuals:
##    Min      1Q  Median      3Q     Max
## -95.42 -31.20  -2.56  27.32 121.61
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 259.898013   3.427300  75.832  < 2e-16 ***
## t            -0.082888   0.009931  -8.346 4.94e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 41.82 on 595 degrees of freedom
## Multiple R-squared:  0.1048, Adjusted R-squared:  0.1033
## F-statistic: 69.66 on 1 and 595 DF,  p-value: 4.937e-16
```

The intercept is 259.898013, and the slope is -0.082888. The relationship here is negative. The p-value here is 4.937e-16, which is smaller than 0.05, therefore, the null hypothesis is rejected, and there is a trend in Hydroelectric Power Consumption

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?

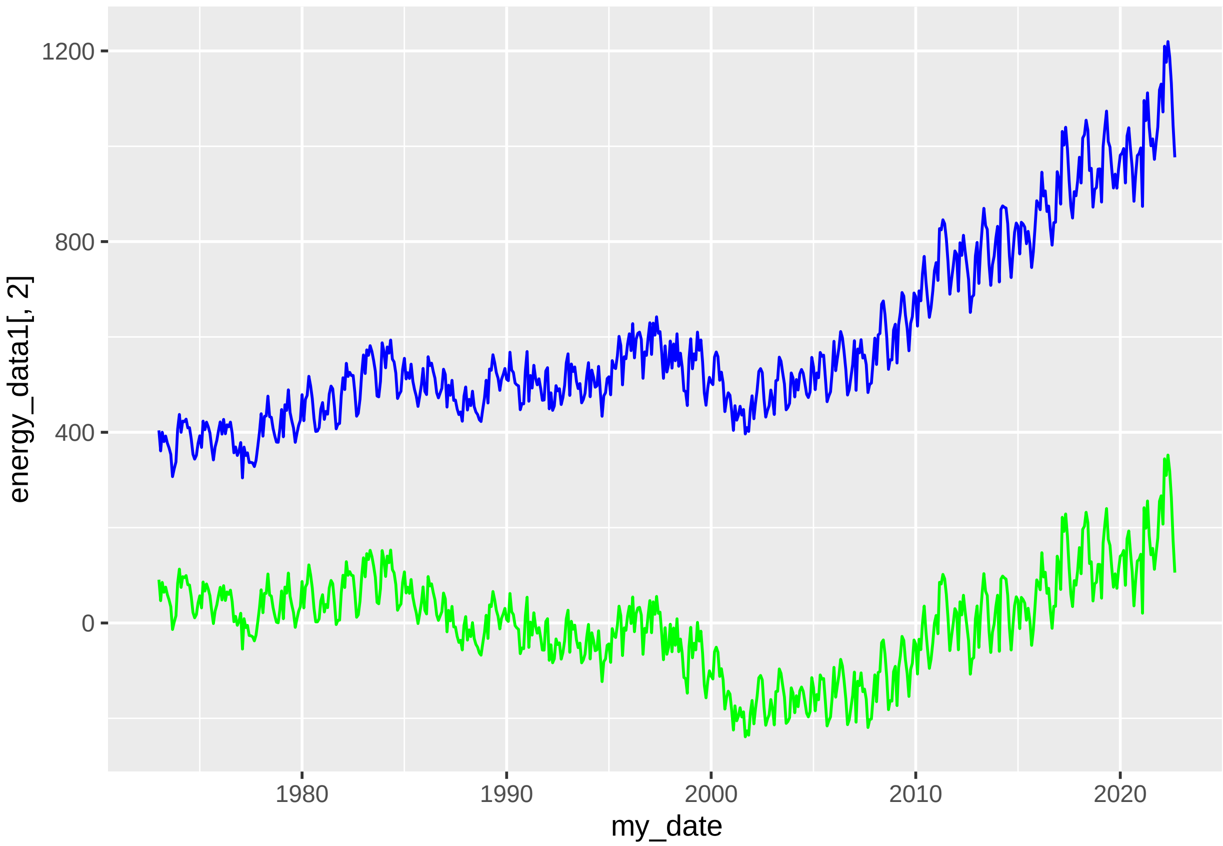
```r
beta0 <- Linear_trend0$coefficients[1]
beta1 <- Linear_trend0$coefficients[2]
detrend_energy_data0 <- energy_data1[,1]-(beta0+beta1*t)

ggplot(energy_data1, aes(x=my_date, y=energy_data1[,1]))+
            geom_line(color="blue")+
            geom_line(aes(y=detrend_energy_data0), col="green")
```




```
          ylab(paste0("Production",colnames(energy_data1)[iHP+1],sep=""))
```

```
## $y
## [1] "ProductionHydroelectric.Power.Consumption"
##
## attr(,"class")
## [1] "labels"
```

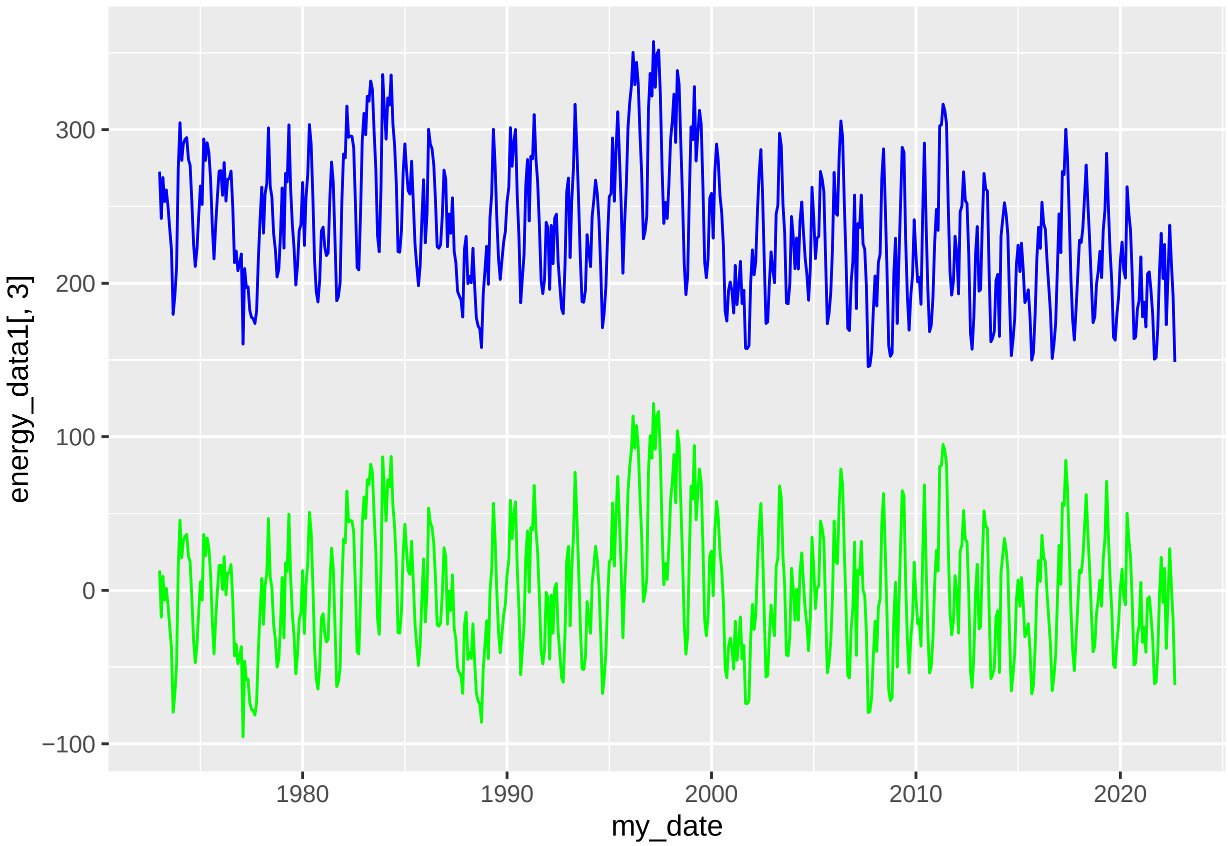
```
beta0 <- Linear_trend1$coefficients[1]
beta1 <- Linear_trend1$coefficients[2]
detrend_energy_data1 <- energy_data1[,2]-(beta0+beta1*t)

ggplot(energy_data1, aes(x=my_date, y=energy_data1[,2]))+
            geom_line(color="blue")+
            geom_line(aes(y=detrend_energy_data1), col="green")
```




```
            ylab(paste0("Production",colnames(energy_data1)[iHP+1],sep=""))
```

```
## $y
## [1] "ProductionHydroelectric.Power.Consumption"
##
## attr(,"class")
## [1] "labels"
```

```
beta0 <- Linear_trend2$coefficients[1]
beta1 <- Linear_trend2$coefficients[2]
detrend_energy_data2 <- energy_data1[,3]-(beta0+beta1*t)

ggplot(energy_data1, aes(x=my_date, y=energy_data1[,3]))+
            geom_line(color="blue")+
            geom_line(aes(y=detrend_energy_data2), col="green")
```




```
            ylab(paste0("Production",colnames(energy_data1)[iHP+1],sep=""))
```

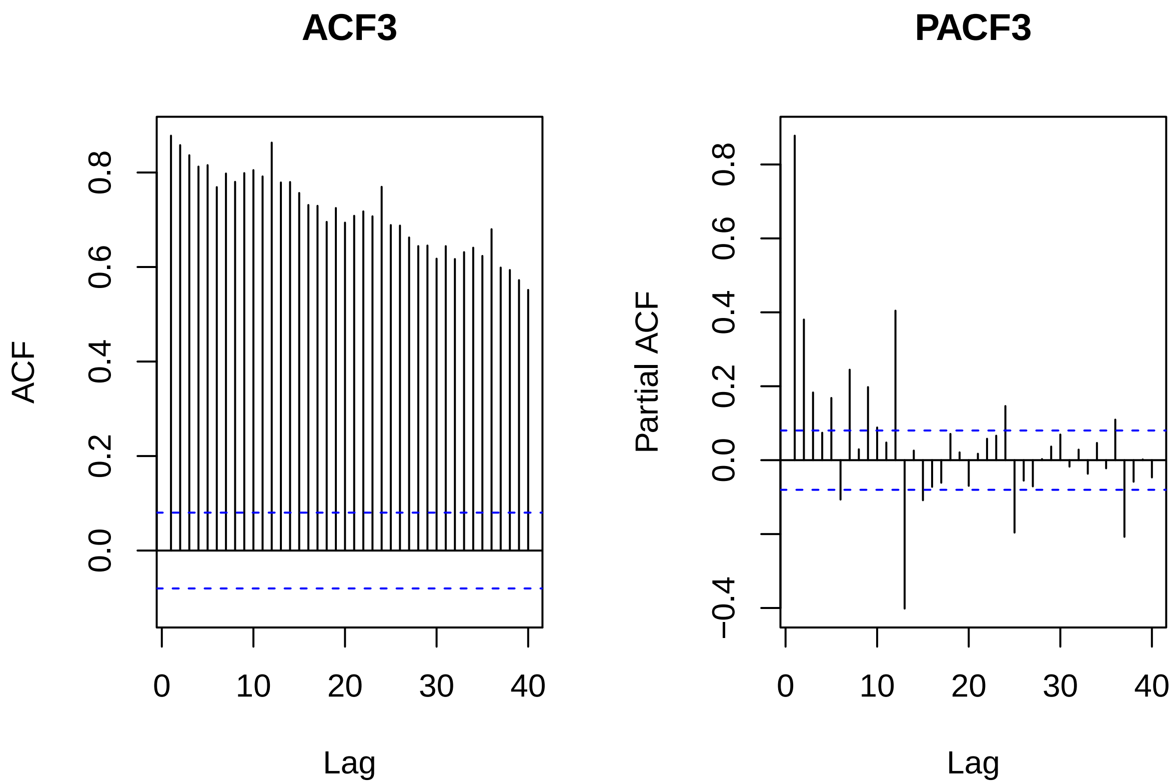
```
## $y
## [1] "ProductionHydroelectric.Power.Consumption"
##
## attr(,"class")
## [1] "labels"
```

After the detrend, first two diagrams lose their trends. They used to have upward trends, but they do not have ang trends now. For the third graph, there is no obvious changes.

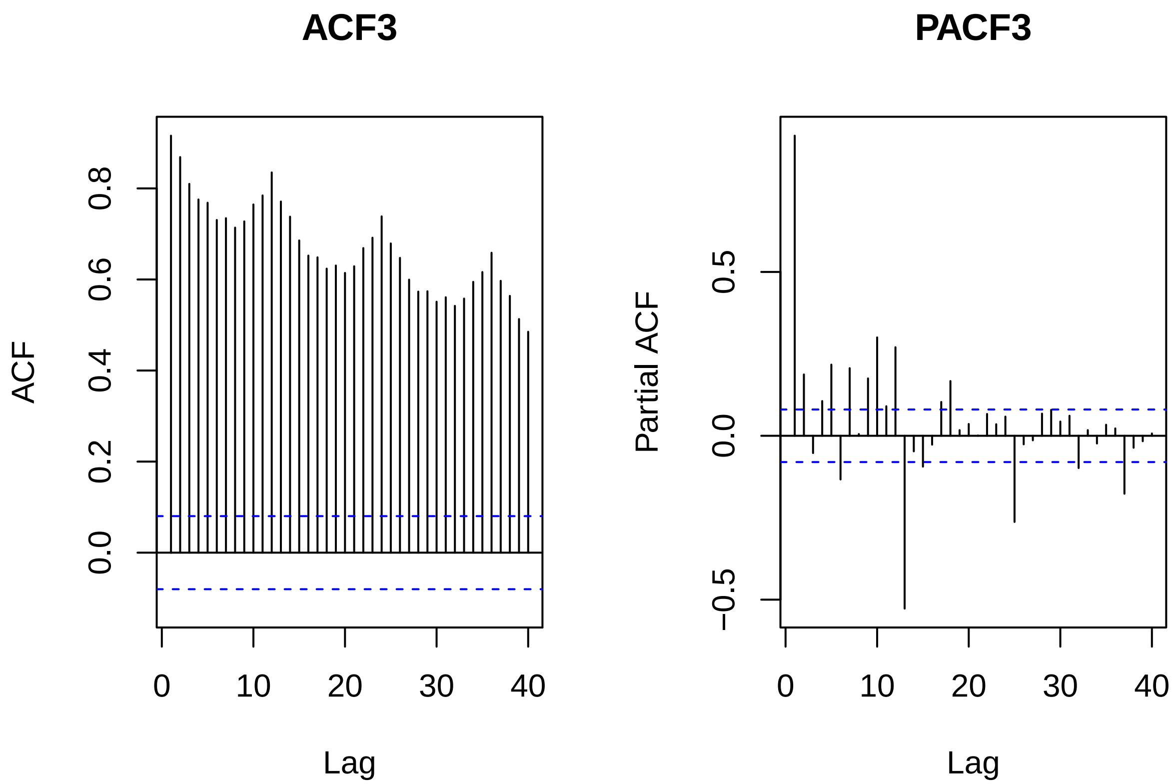
Q5

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

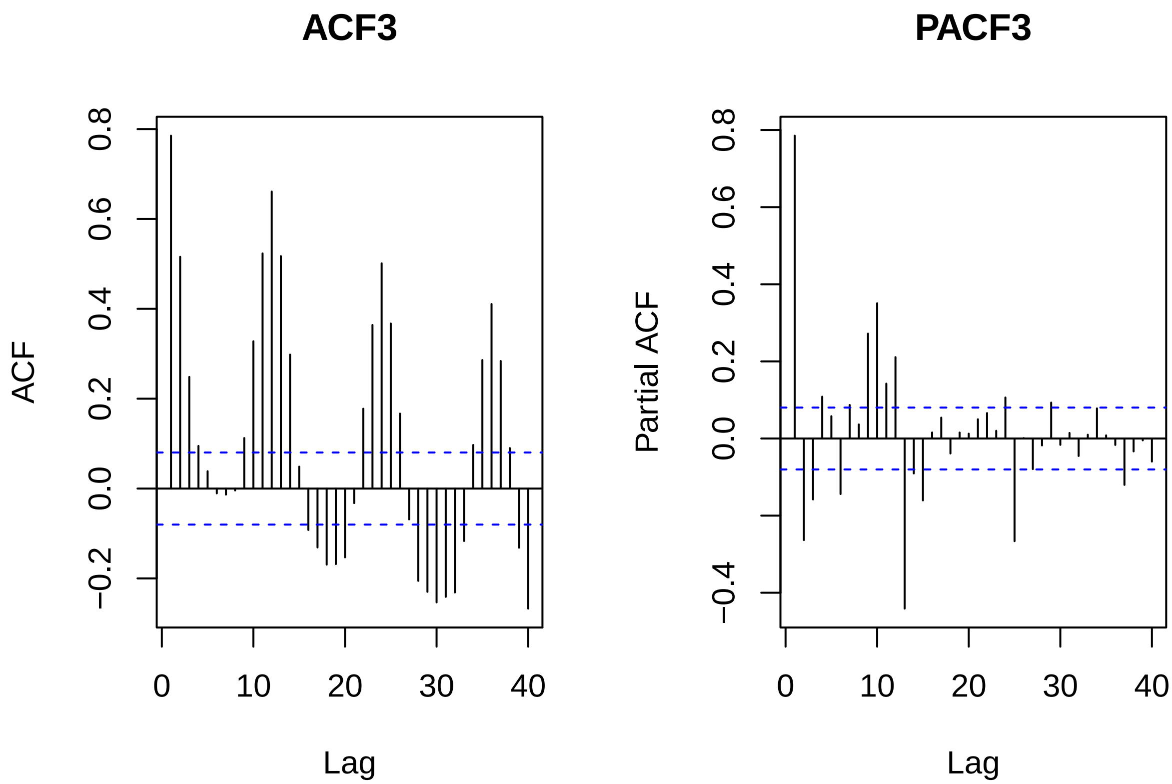
```
ncol_data <- ncol(energy_data1)-1

  par(mfrow=c(1,2))
  Acf(detrend_energy_data0,lag.max=40,main=paste0("ACF",i))
  Pacf(detrend_energy_data0,lag.max=40,main=paste0("PACF",i))
```




```
ncol_data <- ncol(energy_data1)-1

  par(mfrow=c(1,2))
  Acf(detrend_energy_data1,lag.max=40,main=paste0("ACF",i))
  Pacf(detrend_energy_data1,lag.max=40,main=paste0("PACF",i))
```




```
ncol_data <- ncol(energy_data1)-1

  par(mfrow=c(1,2))
  Acf(detrend_energy_data2,lag.max=40,main=paste0("ACF",i))
  Pacf(detrend_energy_data2,lag.max=40,main=paste0("PACF",i))
```



There are not very obvious changes.

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.

```
iHP_0=1
dummies_0 <- seasonaldummy(ts_energy_data1[,iHP_0])
seas_means_model_Biomass=lm(energy_data1[,(iHP_0)]~dummies_0)

summary(seas_means_model_Biomass)
```

```
##
## Call:
## lm(formula = energy_data1[, (iHP_0)] ~ dummies_0)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -160.74  -53.67  -24.36   90.73  181.34
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    288.020     13.163  21.881   <2e-16 ***
## dummies_0Jan    -1.793     18.522  -0.097   0.9229
## dummies_0Feb   -31.102     18.522  -1.679   0.0936 .
```

```
## dummies_0Mar   -9.104     18.522  -0.492   0.6232
## dummies_0Apr  -21.502     18.522  -1.161   0.2462
## dummies_0May  -14.238     18.522  -0.769   0.4424
## dummies_0Jun  -19.602     18.522  -1.058   0.2904
## dummies_0Jul   -3.674     18.522  -0.198   0.8428
## dummies_0Aug   -0.612     18.522  -0.033   0.9737
## dummies_0Sep  -13.335     18.522  -0.720   0.4718
## dummies_0Oct   -4.030     18.615  -0.216   0.8287
## dummies_0Nov   -9.849     18.615  -0.529   0.5970
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 92.14 on 585 degrees of freedom
## Multiple R-squared:  0.01018,    Adjusted R-squared:  -0.008437
## F-statistic: 0.5467 on 11 and 585 DF,  p-value: 0.8714
```

The intercept is 288.020. The p-value here is 0.8714, which is bigger than 0.05, therefore, there is no need for deseason in Biomass Energy production.

```
iHP_1=1
dummies_1 <- seasonaldummy(ts_energy_data1[,iHP_1])
seas_means_model_Renewable=lm(energy_data1[,(iHP_1+1)]~dummies_1)

summary(seas_means_model_Renewable)
```

```
##
## Call:
## lm(formula = energy_data1[, (iHP_1 + 1)] ~ dummies_1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -284.92 -122.23  -68.42   91.22  585.68
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   601.022     27.260  22.048   <2e-16 ***
## dummies_1Jan   11.468     38.358   0.299    0.765
## dummies_1Feb  -41.456     38.358  -1.081    0.280
## dummies_1Mar   23.130     38.358   0.603    0.547
## dummies_1Apr    9.959     38.358   0.260    0.795
## dummies_1May   38.853     38.358   1.013    0.312
## dummies_1Jun   20.378     38.358   0.531    0.595
## dummies_1Jul    8.298     38.358   0.216    0.829
## dummies_1Aug  -19.450     38.358  -0.507    0.612
## dummies_1Sep  -63.770     38.358  -1.662    0.097 .
## dummies_1Oct  -52.612     38.551  -1.365    0.173
## dummies_1Nov  -42.537     38.551  -1.103    0.270
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 190.8 on 585 degrees of freedom
## Multiple R-squared:  0.02844,    Adjusted R-squared:  0.01017
## F-statistic: 1.557 on 11 and 585 DF,  p-value: 0.1076
```

The intercept is 601.022. The p-value here is 0.1076, which is bigger than 0.05, therefore, there is no need for deseason in Renewable Energy production.

```r
iHP_2=2
dummies_2 <- seasonaldummy(ts_energy_data1[,iHP_2])
seas_means_model_Consumption=lm(energy_data1[,(iHP_2+1)]~dummies_2)

summary(seas_means_model_Consumption)
```

```
##
## Call:
## lm(formula = energy_data1[, (iHP_2 + 1)] ~ dummies_2)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -88.99 -23.47  -2.81  21.99 100.18
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   237.225      4.878  48.634  < 2e-16 ***
## dummies_2Jan   13.594      6.864   1.981  0.04811 *
## dummies_2Feb   -8.254      6.864  -1.203  0.22964
## dummies_2Mar   19.980      6.864   2.911  0.00374 **
## dummies_2Apr   15.649      6.864   2.280  0.02297 *
## dummies_2May   39.210      6.864   5.713 1.77e-08 ***
## dummies_2Jun   31.209      6.864   4.547 6.61e-06 ***
## dummies_2Jul   10.436      6.864   1.520  0.12895
## dummies_2Aug  -17.909      6.864  -2.609  0.00931 **
## dummies_2Sep  -50.173      6.864  -7.310 8.82e-13 ***
## dummies_2Oct  -48.262      6.898  -6.996 7.22e-12 ***
## dummies_2Nov  -32.285      6.898  -4.680 3.56e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 34.14 on 585 degrees of freedom
## Multiple R-squared:  0.4132, Adjusted R-squared:  0.4022
## F-statistic: 37.45 on 11 and 585 DF,  p-value: < 2.2e-16
```

The intercept is 237.225. The p-value here less than 0.05, therefore, it is significant and there is need for deseason in Hydroelectric Power Consumption.

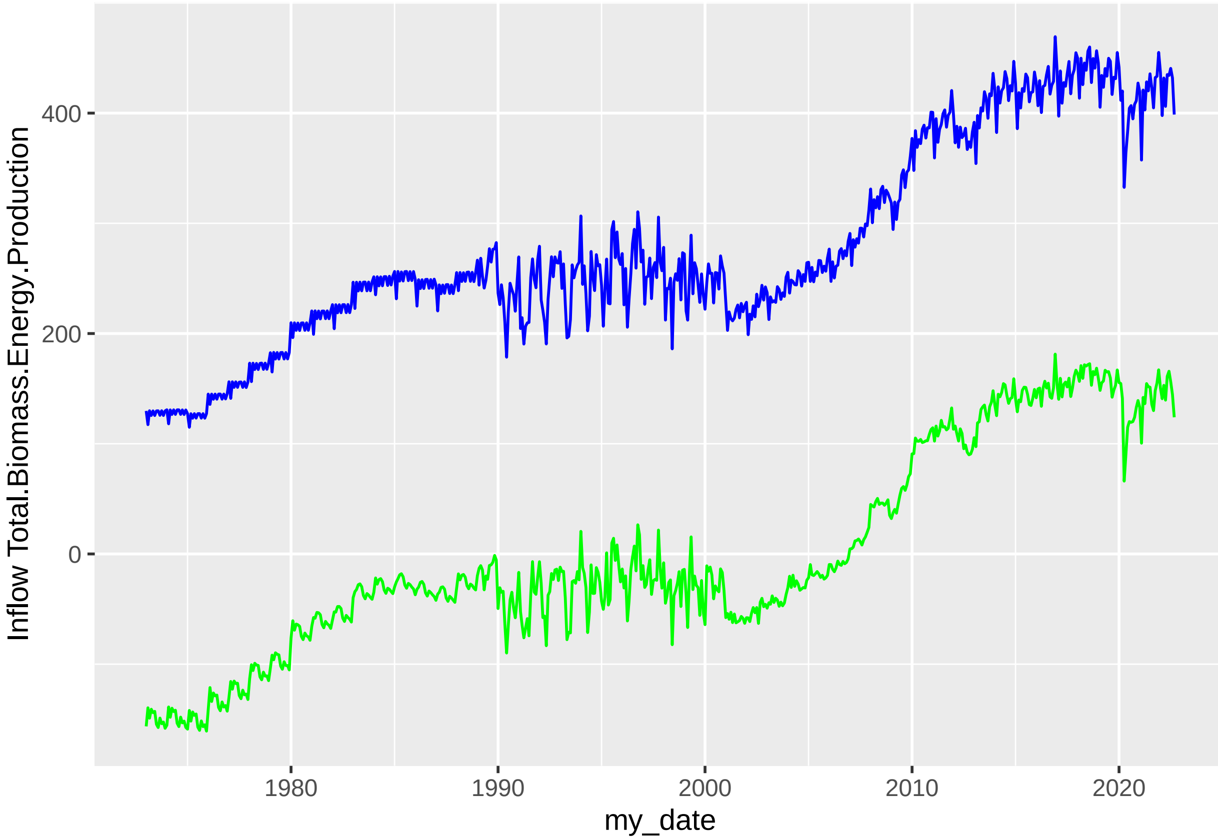
```r
beta_int_Biomass=seas_means_model_Biomass$coefficients[1]
beta_coeff_Biomass=seas_means_model_Biomass$coefficients[2:12]

beta_int_Renewable=seas_means_model_Renewable$coefficients[1]
beta_coeff_Renewable=seas_means_model_Renewable$coefficients[2:12]

beta_int_Consumption=seas_means_model_Consumption$coefficients[1]
beta_coeff_Consumption=seas_means_model_Consumption$coefficients[2:12]
```

Q7

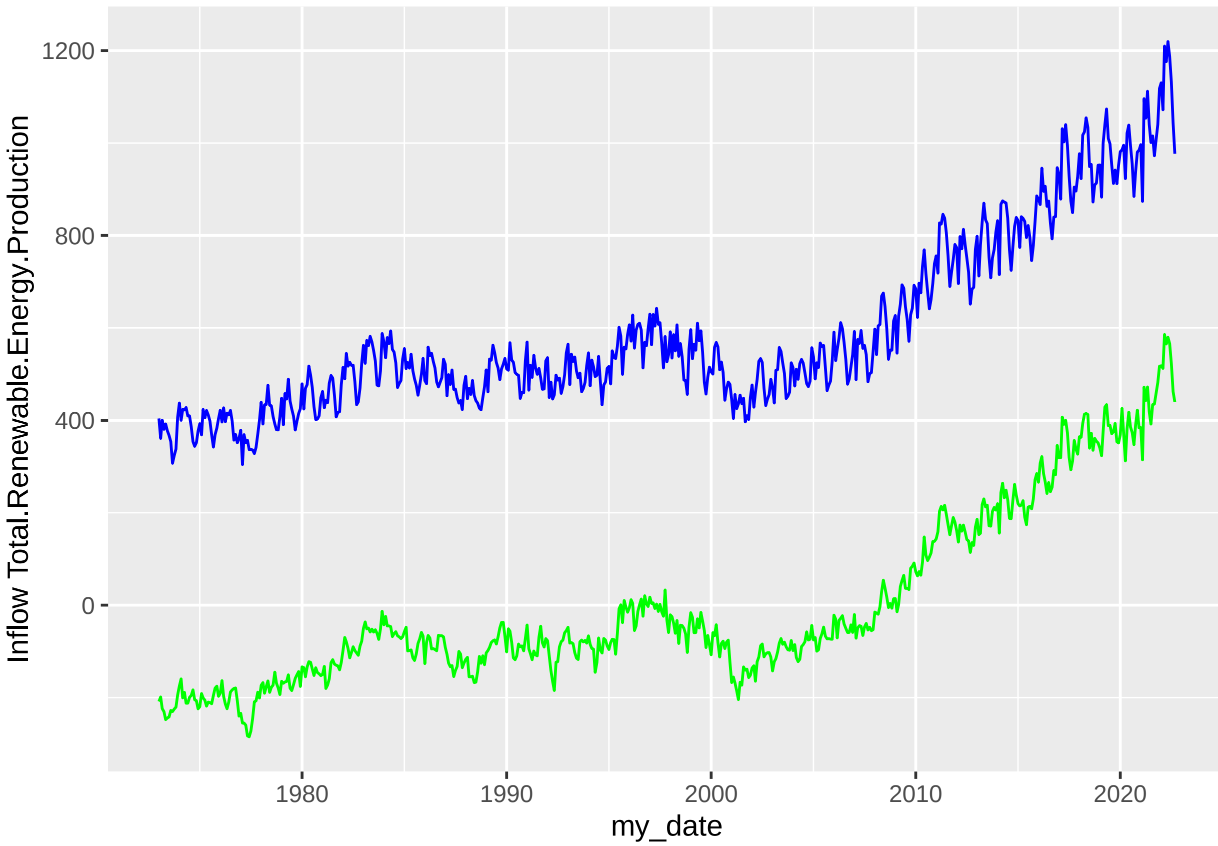
Use the regression coefficients from Q6 to deseason the series. Plot the deseason series and compare with the plots from part Q1. Did anything change?

```r
energy_Biomass_seas_comp=array(0,nobs)
for(i in 1:nobs){
  energy_Biomass_seas_comp[i]=(beta_int_Biomass+beta_coeff_Biomass%*%dummies_0[i,])
```

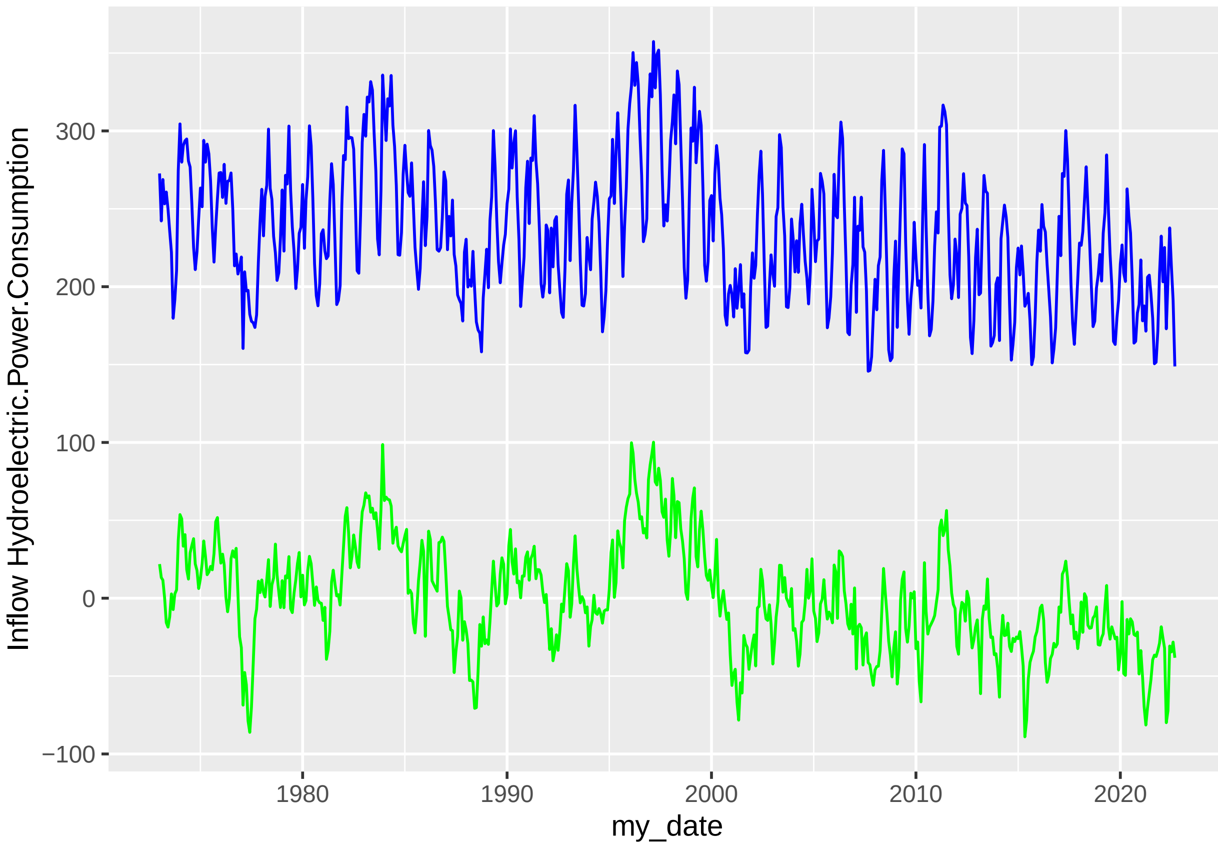
```
}

deseason_energy_Biomass_seas_comp <- energy_data1[,(iHP_1)]-energy_Biomass_seas_comp

ggplot(energy_data1, aes(x=my_date, y=energy_data1[,(iHP_1)])) +
            geom_line(color="blue") +
            ylab(paste0("Inflow ",colnames(energy_data1)[(iHP_1)],sep="")) +
            geom_line(aes(y=deseason_energy_Biomass_seas_comp), col="green")
```




```
energy_Renewable_seas_comp=array(0,nobs)
for(i in 1:nobs){
  energy_Renewable_seas_comp[i]=(beta_int_Renewable+beta_coeff_Renewable%*%dummies_1[i,])
}

deseason_energy_Renewable_seas_comp <- energy_data1[,(1+iHP_1)]-energy_Renewable_seas_comp

ggplot(energy_data1, aes(x=my_date, y=energy_data1[,(1+iHP_1)])) +
            geom_line(color="blue") +
            ylab(paste0("Inflow ",colnames(energy_data1)[(1+iHP_1)],sep="")) +
            geom_line(aes(y=deseason_energy_Renewable_seas_comp ), col="green")
```




```
energy_Consumption_seas_comp=array(0,nobs)
for(i in 1:nobs){
  energy_Consumption_seas_comp[i]=(beta_int_Consumption+beta_coeff_Consumption%*%dummies_2[i,])
}

deseason_energy_Consumption_seas_comp <- energy_data1[,(1+iHP_2)]-energy_Consumption_seas_comp

ggplot(energy_data1, aes(x=my_date, y=energy_data1[,(1+iHP_2)])) +
            geom_line(color="blue") +
            ylab(paste0("Inflow ",colnames(energy_data1)[(1+iHP_2)],sep="")) +
            geom_line(aes(y=deseason_energy_Consumption_seas_comp ), col="green")
```

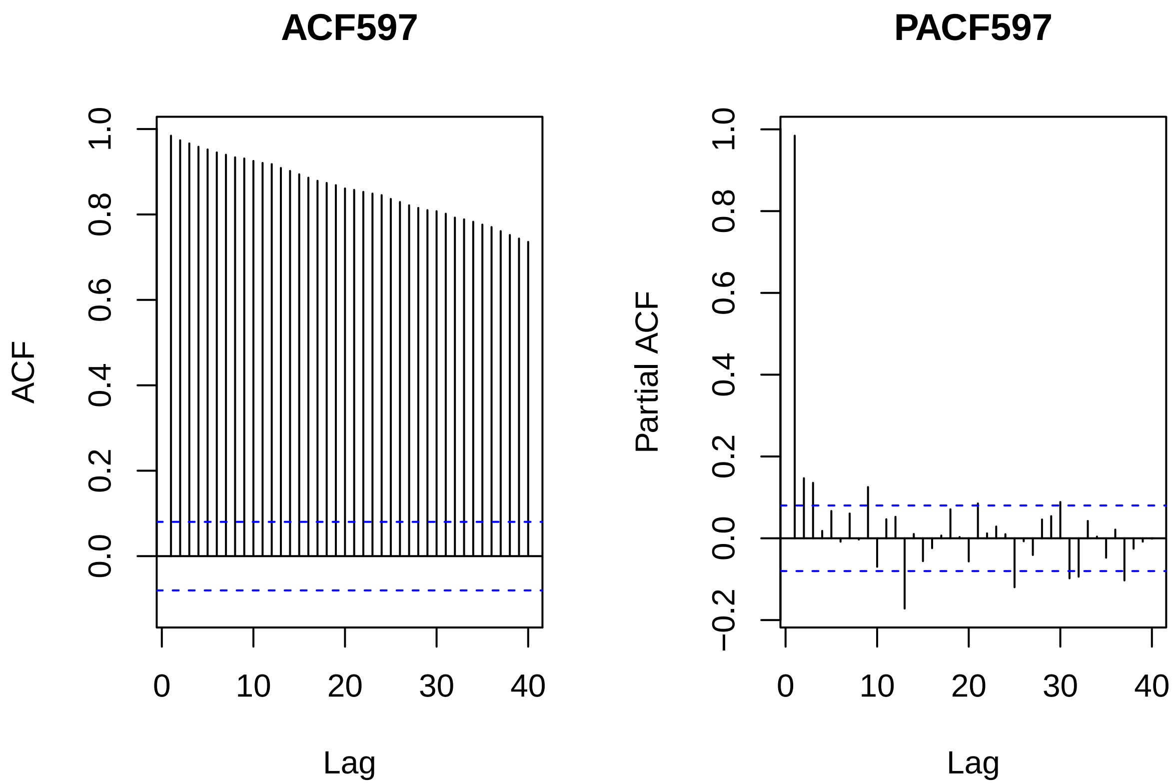


After the deseasoning, there are less fluctuation for all three graphs. Especially for the third one(hydroelectric), it is more obvious than the first two.

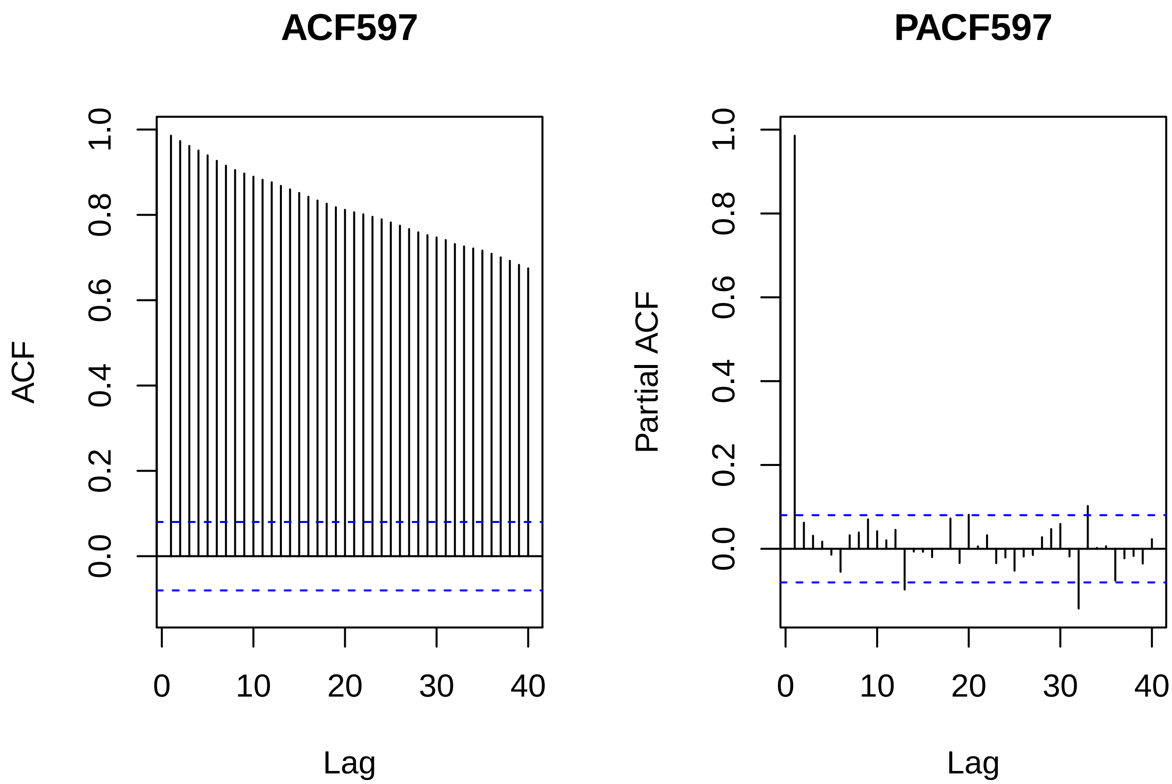
Q8

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

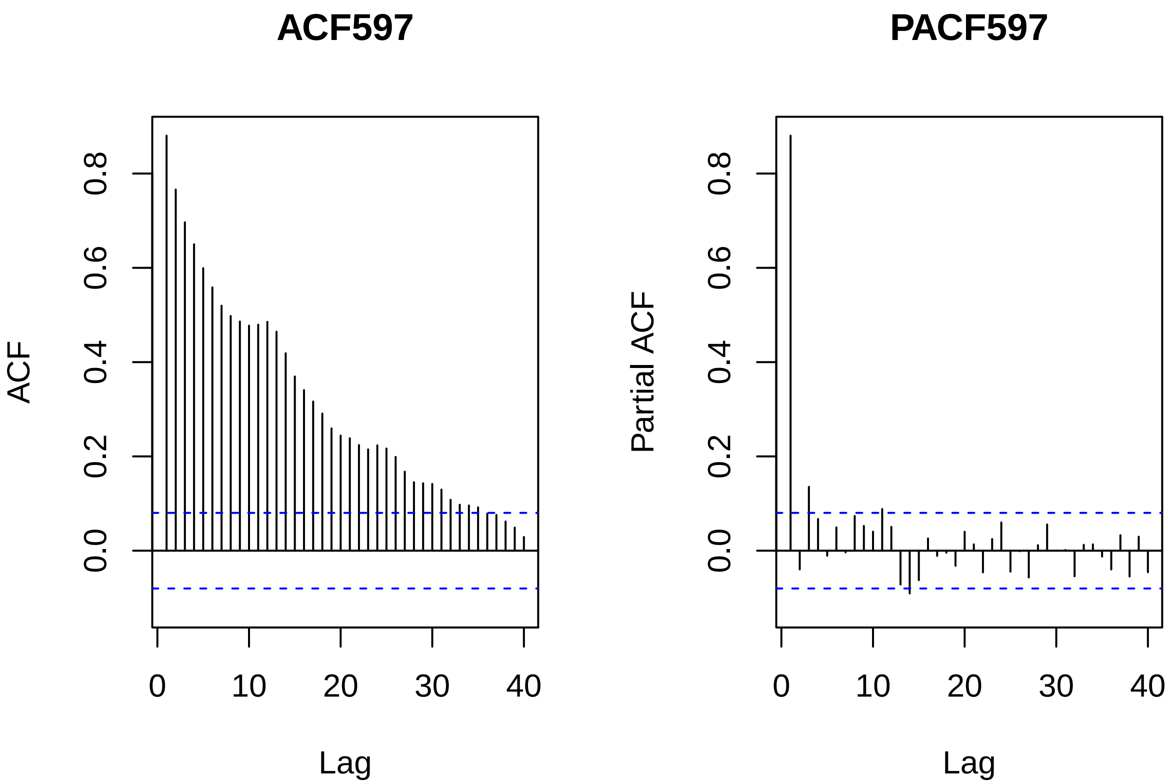
```
ncol_data <- ncol(energy_data1)-1

  par(mfrow=c(1,2))
  Acf(deseason_energy_Biomass_seas_comp,lag.max=40,main=paste0("ACF",i))
  Pacf(deseason_energy_Biomass_seas_comp,lag.max=40,main=paste0("PACF",i))
```




```r
ncol_data <- ncol(energy_data1)-1

  par(mfrow=c(1,2))
  Acf(deseason_energy_Renewable_seas_comp,lag.max=40,main=paste0("ACF",i))
  Pacf(deseason_energy_Renewable_seas_comp,lag.max=40,main=paste0("PACF",i))
```




```
ncol_data <- ncol(energy_data1)-1

  par(mfrow=c(1,2))
  Acf(deseason_energy_Consumption_seas_comp,lag.max=40,main=paste0("ACF",i))
  Pacf(deseason_energy_Consumption_seas_comp,lag.max=40,main=paste0("PACF",i))
```



After the deseasoning, plots of ACF and PACF have big changes. For the ACF, the graph shows a decreasing pattern. For the PACF, only first few lags are out of the dashed line area. In the previous PACF, lots of them are out of the dashed line area. Overall, they are better now.