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

Assignment 3 - Due date 02/08/22

Alex Baad

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_Consumpti 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.\

```
#Load/install required package here
library(lubridate)
```

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

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

```
library(ggplot2)
library(forecast)
```

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

```
library(tseries)
library(Kendall)
library(readxl)
```

##Trend Component

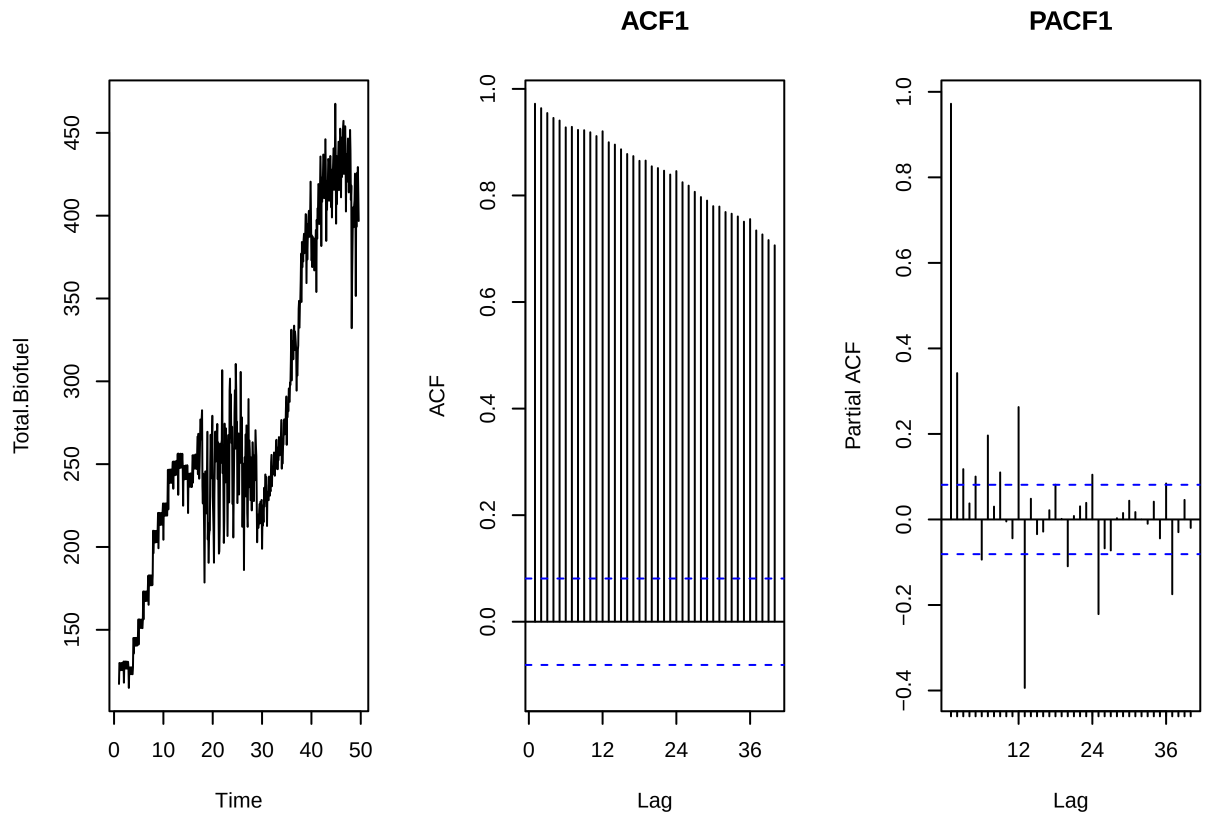
Q1

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)

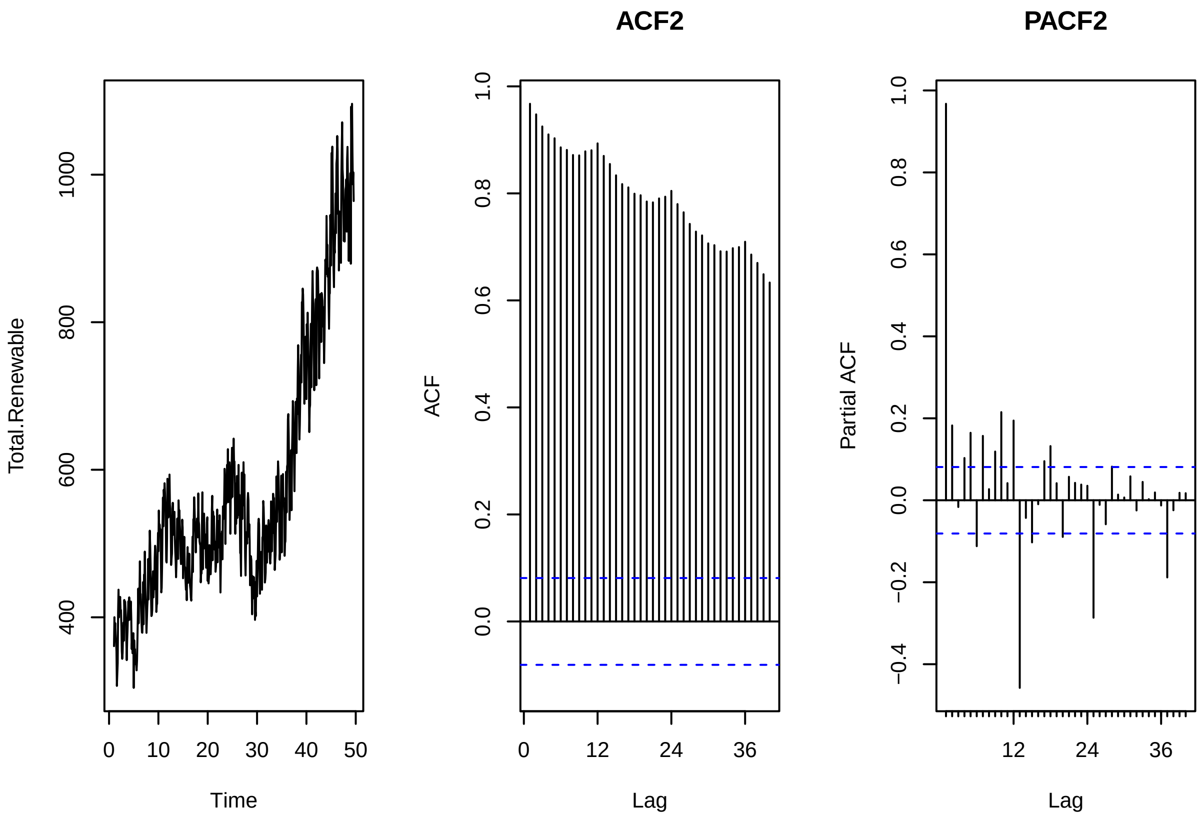
```
energy_data <- read_xlsx(path="./Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source.:
```

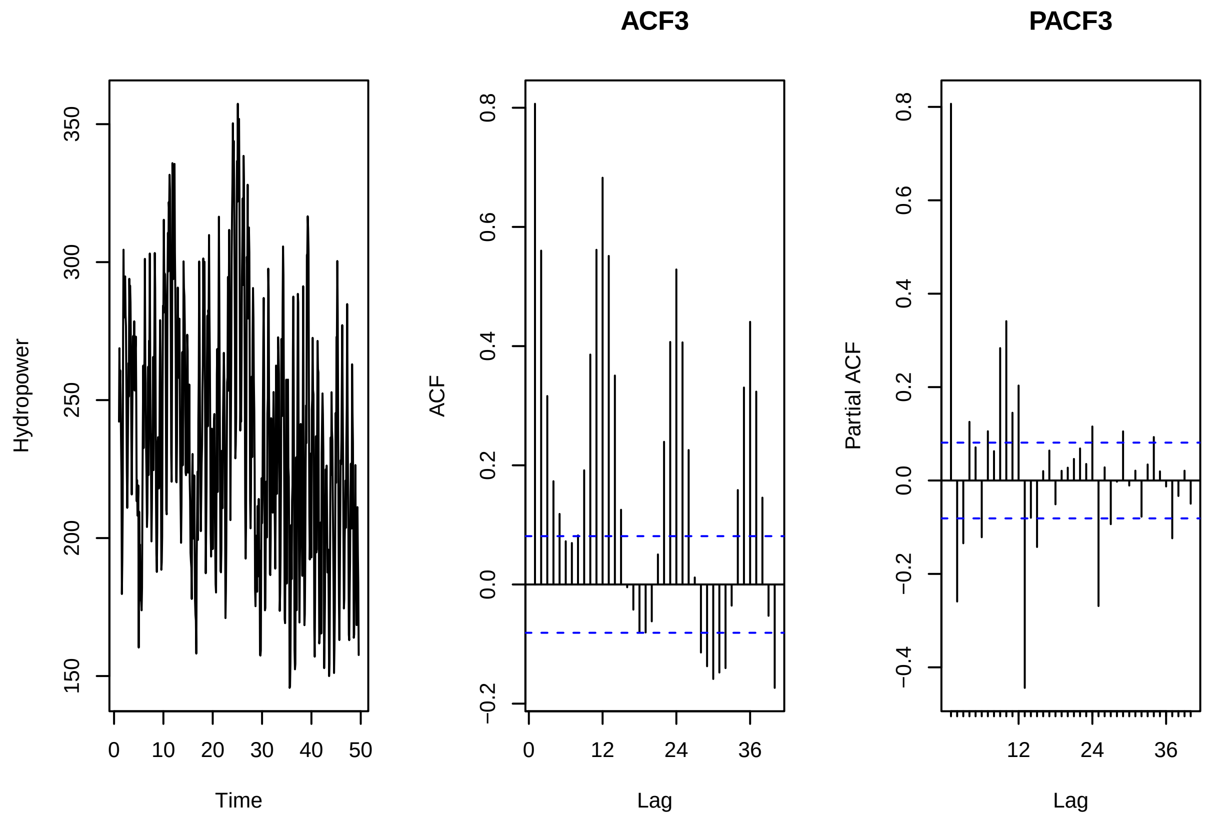
```
## New names:
## * `` -> ...1
## * `` -> ...2
## * `` -> ...3
## * `` -> ...4
## * `` -> ...5
## * ...
```

```
energy_data <- energy_data[-c(1),1:6]
energy_data <- energy_data[,-c(2:3)]
colnames(energy_data)=c("Date","Total.Biofuel","Total.Renewable","Hydropower")
nenergy <- ncol(energy_data)
nobs <- nrow(energy_data)
ts_energy_data <- ts(energy_data[,2:4],frequency=12)

par(mfrow=c(1,3))
for(i in 1:(nenergy-1)){
  ts.plot(ts_energy_data[,i],ylab=colnames(energy_data[,i+1]))
  Acf(ts_energy_data[,i],lag.max=40,main=paste("ACF",i,sep=""))
  Pacf(ts_energy_data[,i],lag.max=40,main=paste("PACF",i,sep=""))
}
```



ACF1
PACF1
Total.Biofuel
Time
ACF
Lag
Partial ACF
Lag


ACF2
PACF2
Total.Renewable
Time
ACF
Lag
Partial ACF
Lag



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?

For Total Biomass Energy Production you can see a general upwards trend based off of the initial plot and looking at the ACF and PACF plot it appears there is some seasonaility as well, though not overly pronounced. For Total Renewable Energy Production it appears to have a general upward trend as well. In the ACF and PACF it appears there is more of a seasonal trend then there was for Total Biomass Energy Production. For Hydroelectric Power Consumption, it doesn't seem to have that much of a trend besides near the end of the data set which appears to have a slightly decreasing trend. For the ACF and PACF, it appears there there is a large seasonal trend, especially compared to Total Biomass Energy Production and Total Renewable Energy Production.

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 the Total Biomass Energy Production, the intercept of 135.3 tells us that at time 0, which aligns with January 1973, it is expected that the total biomass energy production would be 135.3 trillion BTU. The slope of 0.47 for Total Biomass Energy Production tells us that it is expected for an increase of 0.47 trillion Btu each month. For the Total Renewable Energy Production, the intercept of 323.51 tells us that at time 0, which aligns with January 1973, it is expected that the total renewable energy production would be

323.51 trillion BTU. The slope of 0.88 for Total Renewable Energy Production tells us that it is expected for an increase of 0.88 trillion Btu each month. For the Hydroelectric Power Consumption, the intercept of 259.01 tells us that at time 0, which aligns with January 1973, it is expected that the hydroelectric power consumption would be 259.01 trillion BTU. The slope of -0.08 for Hydroelectric Power Consumption tells us that it is expected for a decrease of 0.08 trillion Btu each month.

```
t <- c(1:nobs)

linear_trend_model_TBEP=lm(ts_energy_data[,1]~t)
summary(linear_trend_model_TBEP)
```

```
##
## Call:
## lm(formula = ts_energy_data[, 1] ~ t)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -101.90  -24.40    4.98   33.23   82.31
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.353e+02  3.288e+00   41.16   <2e-16 ***
## t           4.743e-01  9.738e-03   48.70   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 39.67 on 582 degrees of freedom
## Multiple R-squared:  0.803,  Adjusted R-squared:  0.8026
## F-statistic:  2372 on 1 and 582 DF,  p-value: < 2.2e-16
```

```
beta0_TBEP=as.numeric(linear_trend_model_TBEP$coefficients[1])
beta1_TBEP=as.numeric(linear_trend_model_TBEP$coefficients[2])

linear_trend_model_TREP=lm(ts_energy_data[,2]~t)
summary(linear_trend_model_TREP)
```

```
##
## Call:
## lm(formula = ts_energy_data[, 2] ~ t)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -230.424  -58.006    5.735   62.538  261.081
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 323.5141     8.0346   40.27   <2e-16 ***
## t             0.8819     0.0238   37.06   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 96.96 on 582 degrees of freedom
## Multiple R-squared:  0.7023, Adjusted R-squared:  0.7018
## F-statistic:  1373 on 1 and 582 DF,  p-value: < 2.2e-16
```

```
beta0_TREP=as.numeric(linear_trend_model_TREP$coefficients[1])
beta1_TREP=as.numeric(linear_trend_model_TREP$coefficients[2])

linear_trend_model_HPC=lm(ts_energy_data[,3]~t)
summary(linear_trend_model_HPC)
```

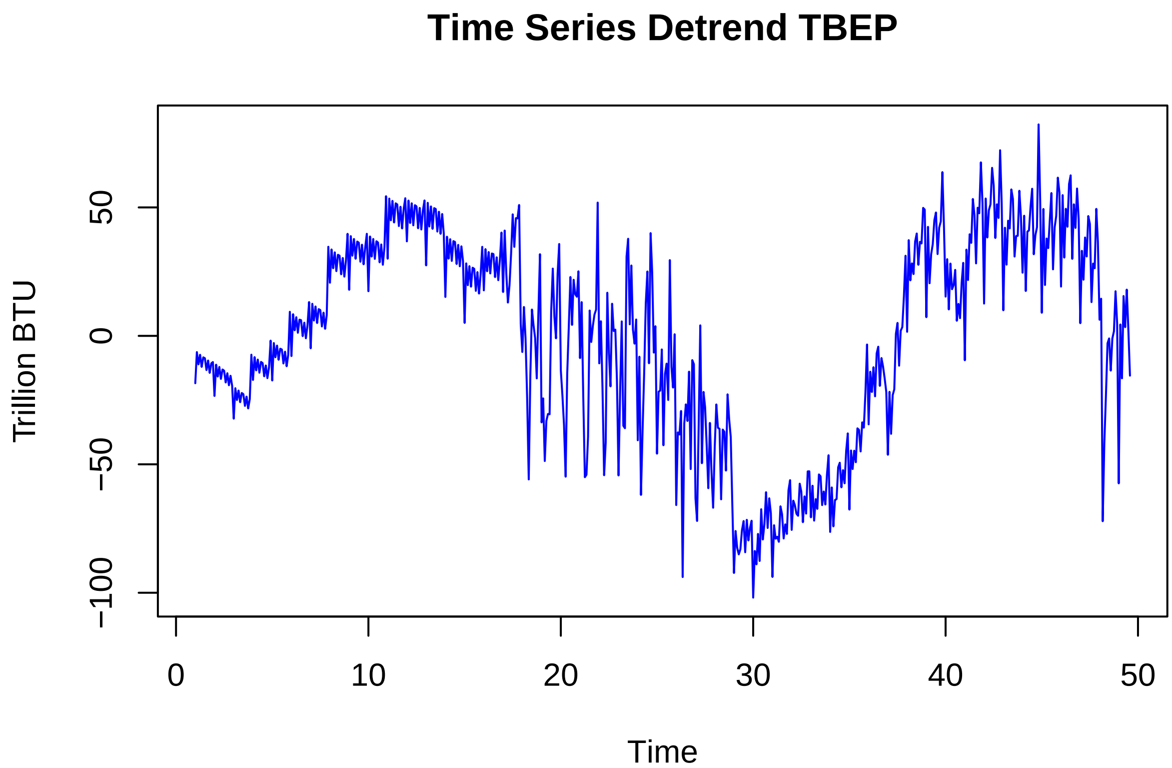
```
##
## Call:
## lm(formula = ts_energy_data[, 3] ~ t)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -94.810 -31.364  -2.446  27.901 121.287
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 259.01040    3.48029  74.422  < 2e-16 ***
## t            -0.07900    0.01031  -7.664 7.61e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 42 on 582 degrees of freedom
## Multiple R-squared:  0.09166,    Adjusted R-squared:  0.0901
## F-statistic: 58.73 on 1 and 582 DF,  p-value: 7.607e-14
```

```
beta0_HPC=as.numeric(linear_trend_model_HPC$coefficients[1])
beta1_HPC=as.numeric(linear_trend_model_HPC$coefficients[2])
```

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? For the Total Biomass Energy Production (TBEP) and Total Renewable Energy Production (TREP), the overall positive trends are now gone, but there appears to still be some seasonal trend for both. For the Hydroelectric Power Consumption (HPC), it appears that not much has changed besides the last part has less of a downward trend. This may be due to the slops for HBEP and HPEP being at least four times as large or greater than the slope trend for HPC which means the change for HPEP and HBEP will be more drastic than for HPC.

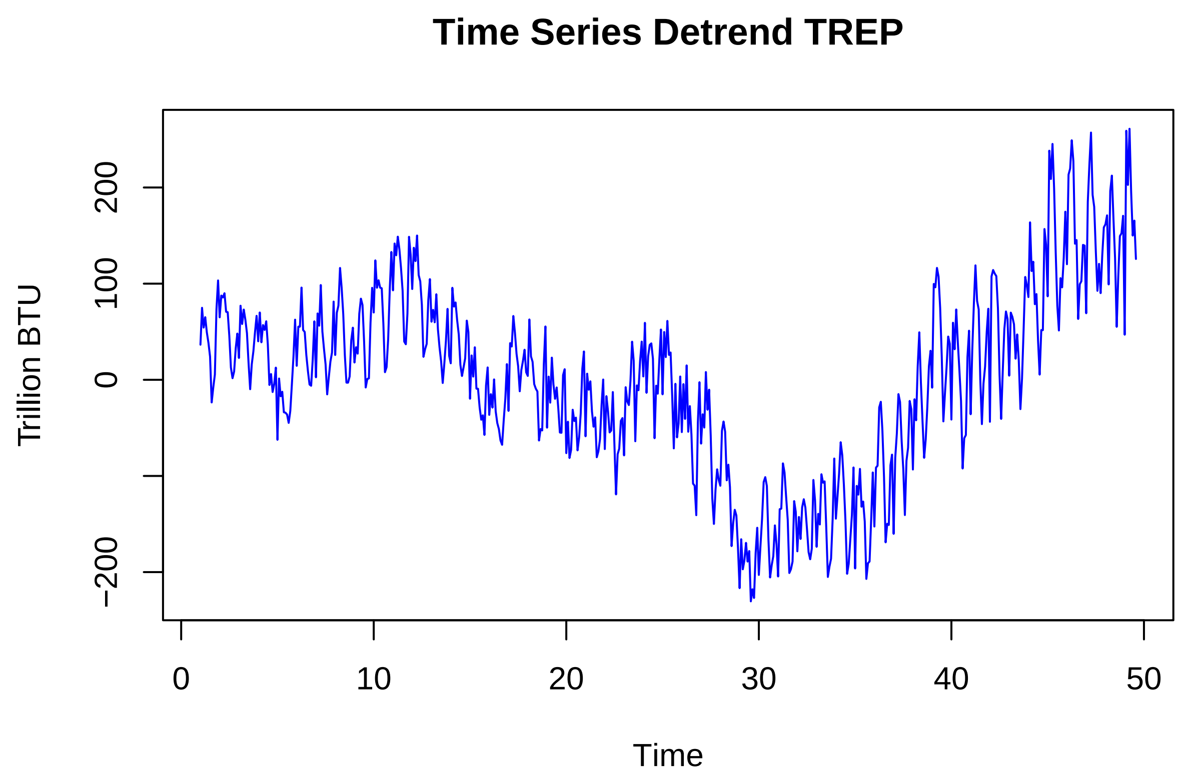
```
detrend_energy_data_TBEP <- ts_energy_data[,1]-(beta0_TBEP+beta1_TBEP*t)
detrend_energy_data_TREP <- ts_energy_data[,2]-(beta0_TREP+beta1_TREP*t)
detrend_energy_data_HPC <- ts_energy_data[,3]-(beta0_HPC+beta1_HPC*t)

plot(detrend_energy_data_TBEP,col="blue",ylab="Trillion BTU",main="Time Series Detrend TBEP")
```




```r
geom_abline(intercept = beta0_TBEP, slope = beta1_TBEP, color="red")
```

```
## mapping: intercept = ~intercept, slope = ~slope
## geom_abline: na.rm = FALSE
## stat_identity: na.rm = FALSE
## position_identity
```

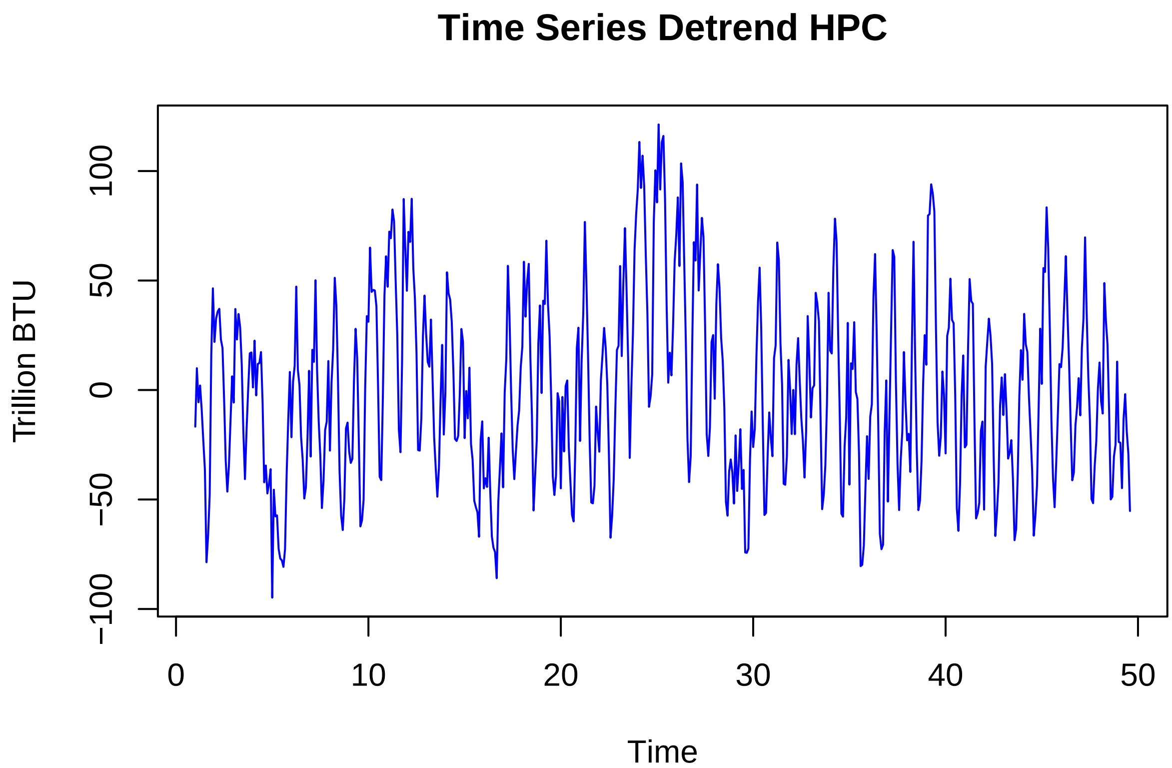
```r
plot(detrend_energy_data_TREP,col="blue",ylab="Trillion BTU",main="Time Series Detrend TREP")
```




```
geom_abline(intercept = beta0_TREP, slope = beta1_TREP, color="red")
```

```
## mapping: intercept = ~intercept, slope = ~slope
## geom_abline: na.rm = FALSE
## stat_identity: na.rm = FALSE
## position_identity
```

```
plot(detrend_energy_data_HPC,col="blue",ylab="Trillion BTU",main="Time Series Detrend HPC")
```



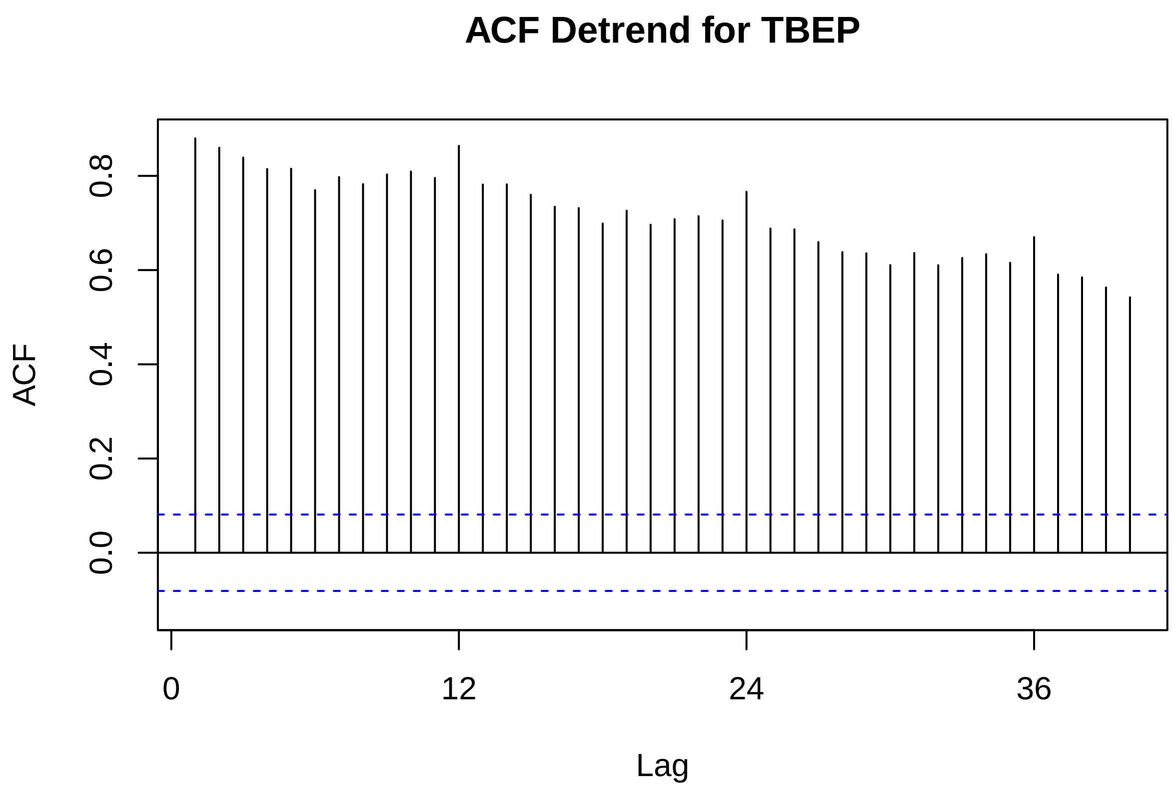

```
geom_abline(intercept = beta0_HPC, slope = beta1_HPC, color="red")
```

```
## mapping: intercept = ~intercept, slope = ~slope
## geom_abline: na.rm = FALSE
## stat_identity: na.rm = FALSE
## position_identity
```

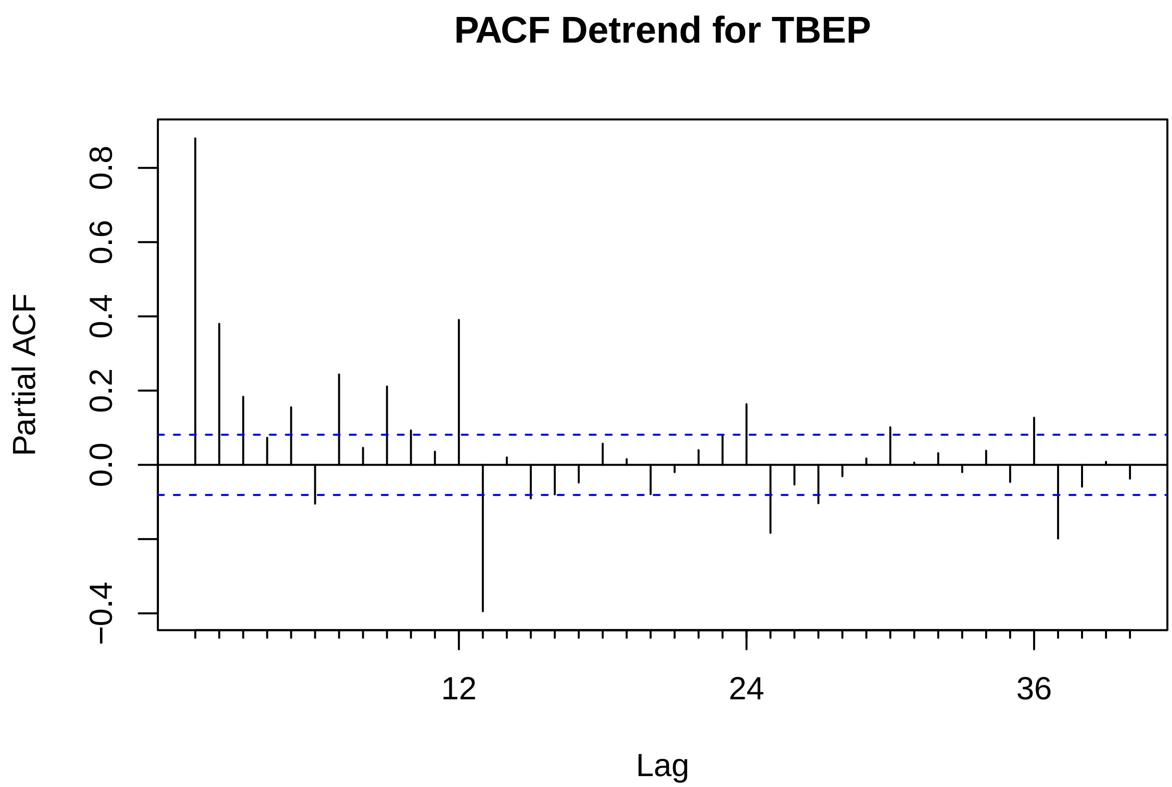
Q5

Plot ACF and PACF for the detrended series and compare with the plots from Q1. Did the plots change? How? Yes the ACF plots for all three seemed to have more of a pronounced seasonal trend. The PACF plots for all three seem to have less variability from before.

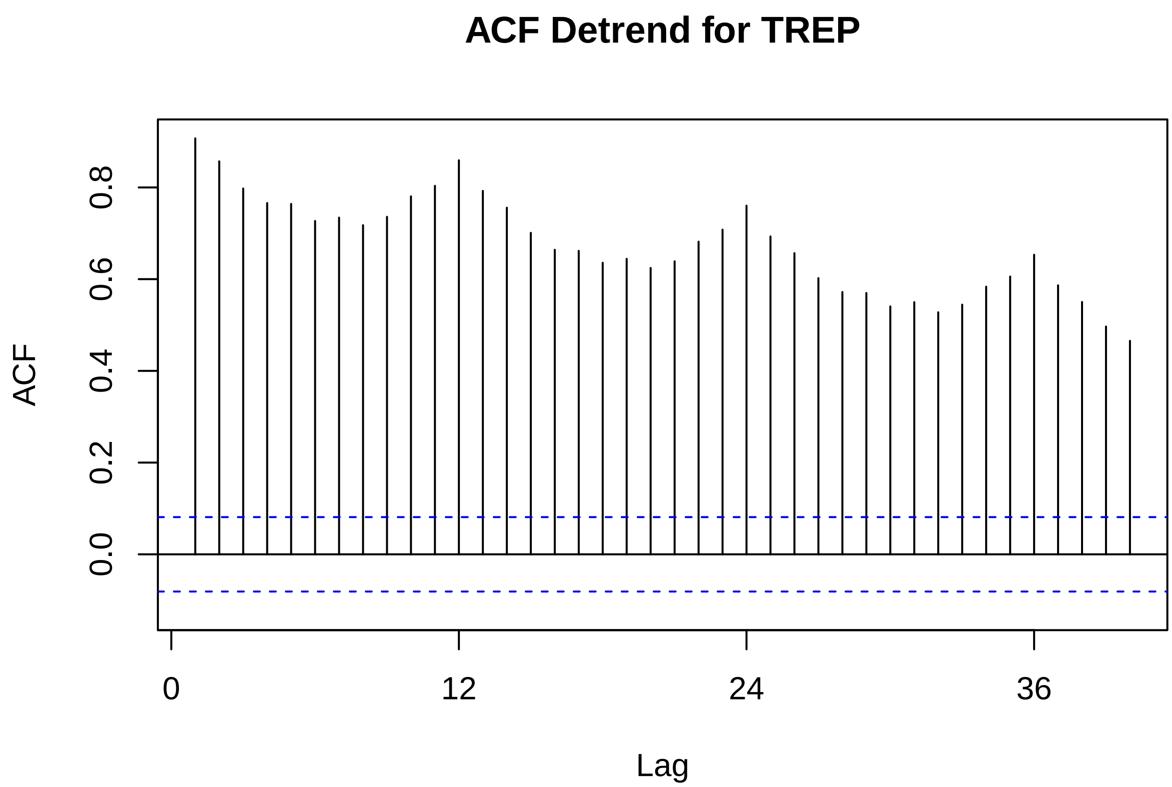
```
Acf(detrend_energy_data_TBEP,lag.max=40,main="ACF Detrend for TBEP")
```



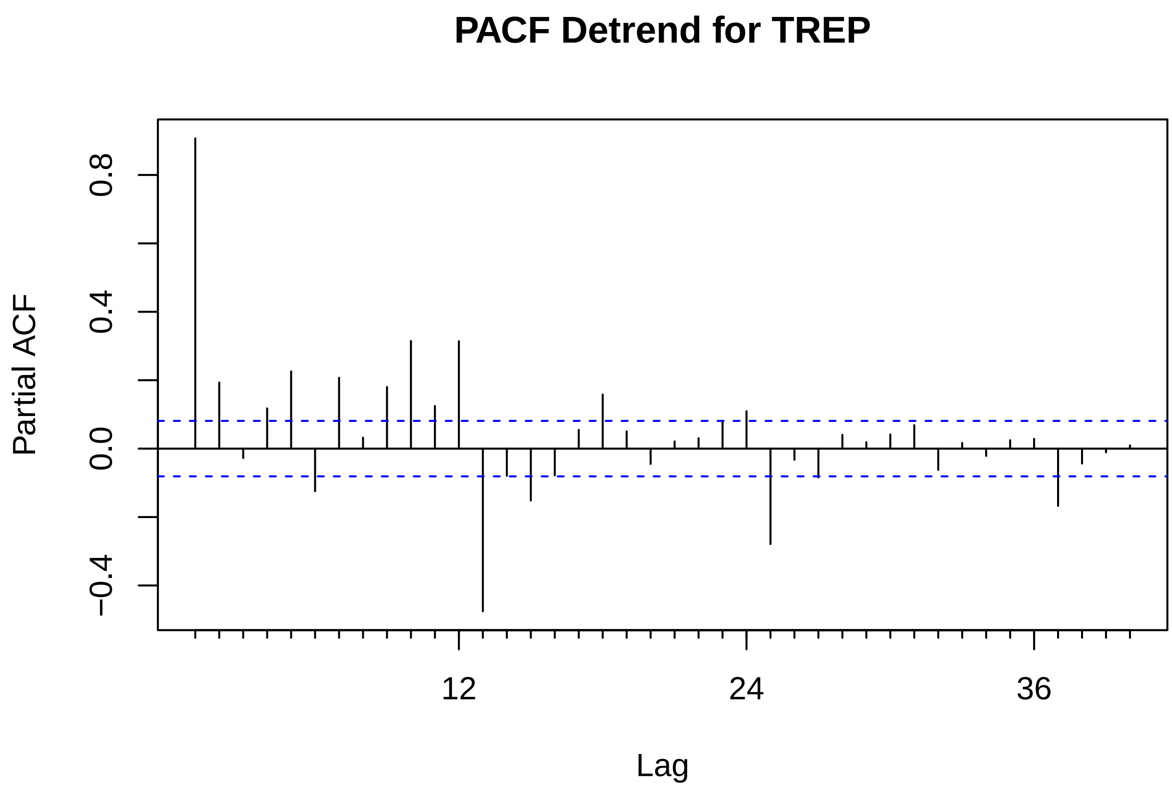

```
Pacf(detrend_energy_data_TBEP,lag.max=40,main="PACF Detrend for TBEP")
```



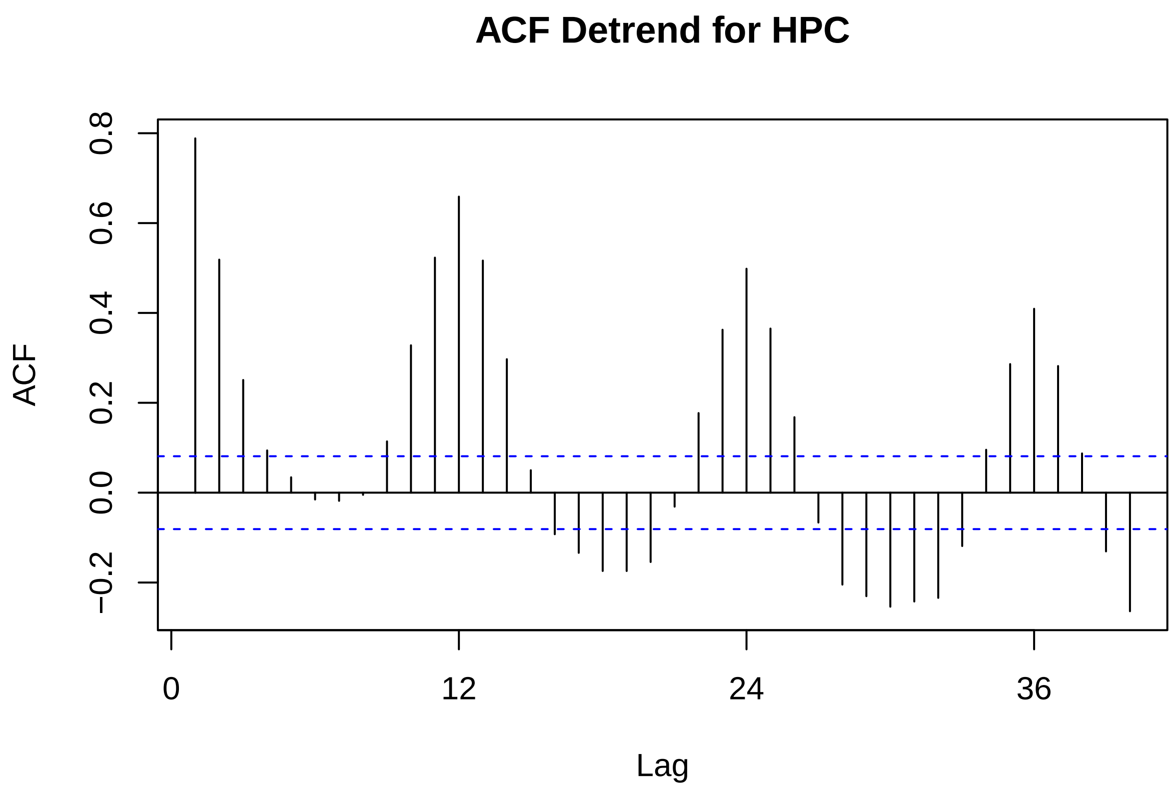

```
Acf(detrend_energy_data_TREP,lag.max=40,main="ACF Detrend for TREP")
```



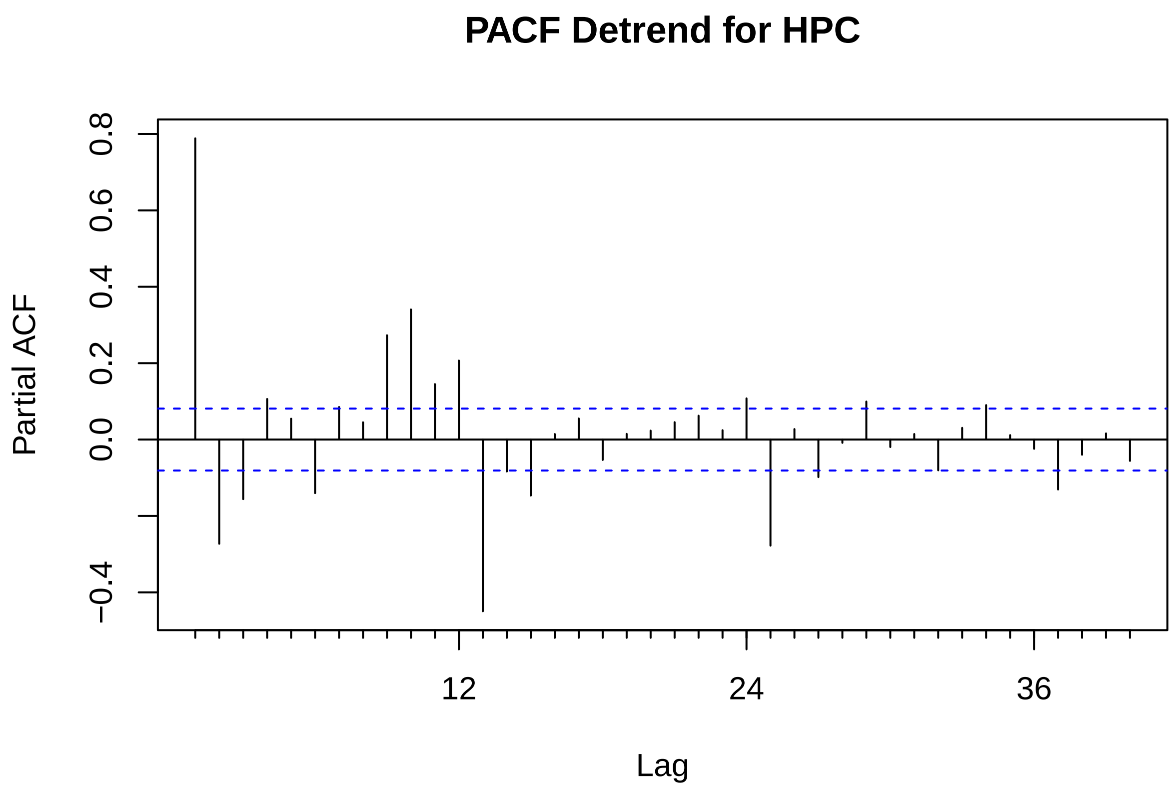

```
Pacf(detrend_energy_data_TREP,lag.max=40,main="PACF Detrend for TREP")
```




```
Acf(detrend_energy_data_HPC,lag.max=40,main="ACF Detrend for HPC")
```




```
Pacf(detrend_energy_data_HPC,lag.max=40,main="PACF Detrend for HPC")
```



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. After running the seasonal trend test for TBEP, it appears that there is no seasonal trend since the final p value, which measures the expected difference in trillion BTU from a given december to each month is not statisitcally significant. This measures whether the null hypothesis, which is that there is a seasonal trend, can be rejected. The same can be said for TREP where the final p value is also not statistically significant. However, for HPC, there is statistically significant evidence for every month besides February, March and June and the final p value that there is statistically significant evidence of a seasonal trend.

```
dummies <- seasonaldummy(ts_energy_data[,1])
seas_means_model_TBEP=lm(energy_data$Total.Biofuel~dummies)
summary(seas_means_model_TBEP)
```

```
##
## Call:
## lm(formula = energy_data$Total.Biofuel ~ dummies)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -158.66   -50.96   -21.95    61.26   183.31
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  285.930     12.940  22.097   <2e-16 ***
## dummiesJan   -32.270     18.206  -1.772   0.0768 .
## dummiesFeb   -10.561     18.206  -0.580   0.5621
## dummiesMar   -22.697     18.206  -1.247   0.2130
## dummiesApr   -15.753     18.206  -0.865   0.3873
## dummiesMay   -21.289     18.206  -1.169   0.2428
## dummiesJun    -5.187     18.206  -0.285   0.7758
## dummiesJul    -1.940     18.206  -0.107   0.9152
## dummiesAug   -14.206     18.206  -0.780   0.4355
## dummiesSep    -5.317     18.300  -0.291   0.7715
## dummiesOct   -11.281     18.300  -0.616   0.5379
## dummiesNov    -1.688     18.300  -0.092   0.9265
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 89.65 on 572 degrees of freedom
## Multiple R-squared:  0.01127,    Adjusted R-squared:  -0.007742
## F-statistic: 0.5928 on 11 and 572 DF,  p-value: 0.8354
```

```
beta_int_TBEP=seas_means_model_TBEP$coefficients[1]
beta_coeff_TBEP=seas_means_model_TBEP$coefficients[2:12]

dummies <- seasonaldummy(ts_energy_data[,2])
seas_means_model_TREP=lm(energy_data$Total.Renewable~dummies)
summary(seas_means_model_TREP)
```

```
##
## Call:
## lm(formula = energy_data$Total.Renewable ~ dummies)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -272.95 -111.22   -59.37    65.76   480.41
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  605.884     25.458  23.799   <2e-16 ***
## dummiesJan   -56.905     35.819  -1.589   0.1127
## dummiesFeb     5.978     35.819   0.167   0.8675
## dummiesMar    -7.005     35.819  -0.196   0.8450
## dummiesApr    21.587     35.819   0.603   0.5470
## dummiesMay     3.552     35.819   0.099   0.9210
## dummiesJun    -7.798     35.819  -0.218   0.8277
## dummiesJul   -34.272     35.819  -0.957   0.3391
## dummiesAug   -78.028     35.819  -2.178   0.0298 *
## dummiesSep   -67.290     36.003  -1.869   0.0621 .
## dummiesOct   -57.702     36.003  -1.603   0.1095
## dummiesNov   -15.913     36.003  -0.442   0.6587
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 176.4 on 572 degrees of freedom
## Multiple R-squared:  0.03191,    Adjusted R-squared:  0.01329
## F-statistic: 1.714 on 11 and 572 DF,  p-value: 0.06681
```

```
beta_int_TREP=seas_means_model_TREP$coefficients[1]
beta_coeff_TREP=seas_means_model_TREP$coefficients[2:12]

dummies <- seasonaldummy(ts_energy_data[,3])
seas_means_model_HPC=lm(energy_data$Hydropower~dummies)
summary(seas_means_model_HPC)
```

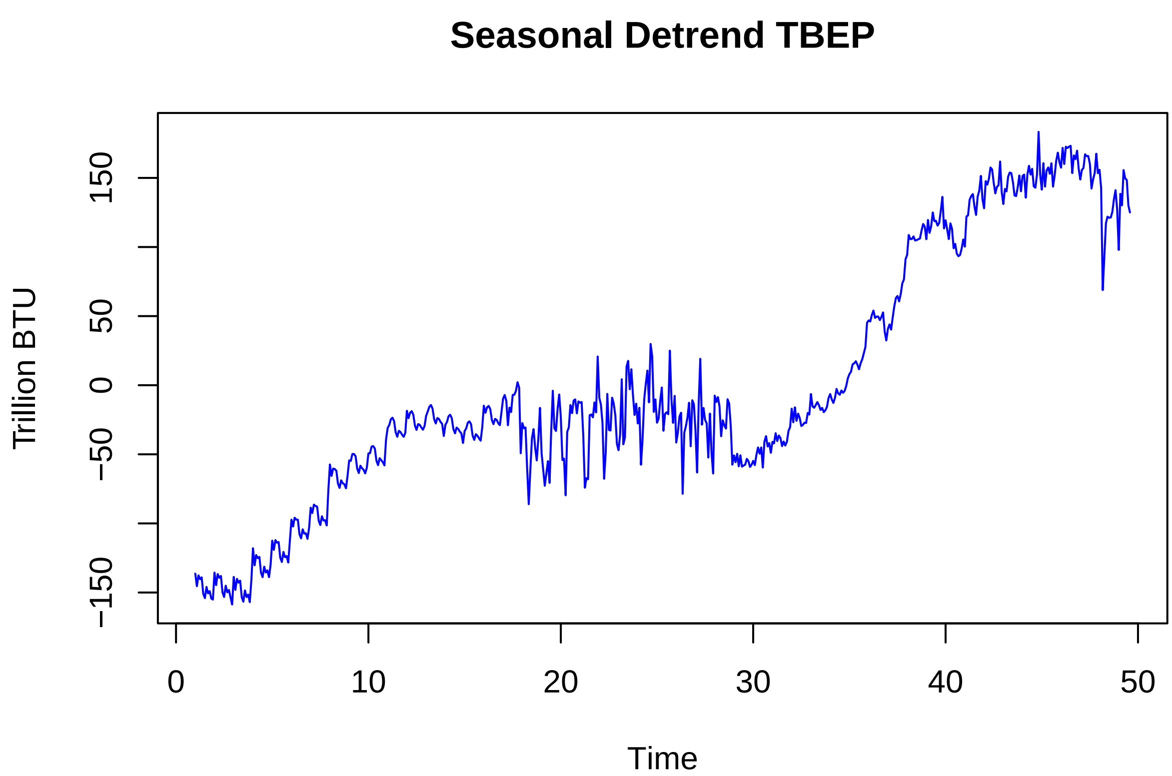
```
##
## Call:
## lm(formula = energy_data$Hydropower ~ dummies)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -90.253 -22.978  -2.995  21.503  99.478
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  250.955      4.895  51.270  < 2e-16 ***
## dummiesJan   -21.204      6.887  -3.079 0.002178 **
## dummiesFeb     6.954      6.887   1.010 0.313066
## dummiesMar     3.505      6.887   0.509 0.610944
## dummiesApr    26.847      6.887   3.898 0.000108 ***
## dummiesMay    18.201      6.887   2.643 0.008446 **
## dummiesJun    -2.603      6.887  -0.378 0.705588
## dummiesJul   -30.967      6.887  -4.496 8.37e-06 ***
## dummiesAug   -62.966      6.887  -9.143  < 2e-16 ***
## dummiesSep   -61.199      6.922  -8.841  < 2e-16 ***
## dummiesOct   -45.301      6.922  -6.544 1.33e-10 ***
## dummiesNov   -13.114      6.922  -1.894 0.058673 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 33.91 on 572 degrees of freedom
## Multiple R-squared:  0.4179, Adjusted R-squared:  0.4068
## F-statistic: 37.34 on 11 and 572 DF,  p-value: < 2.2e-16
```

```
beta_int_HPC=seas_means_model_HPC$coefficients[1]
beta_coeff_HPC=seas_means_model_HPC$coefficients[2:12]
```

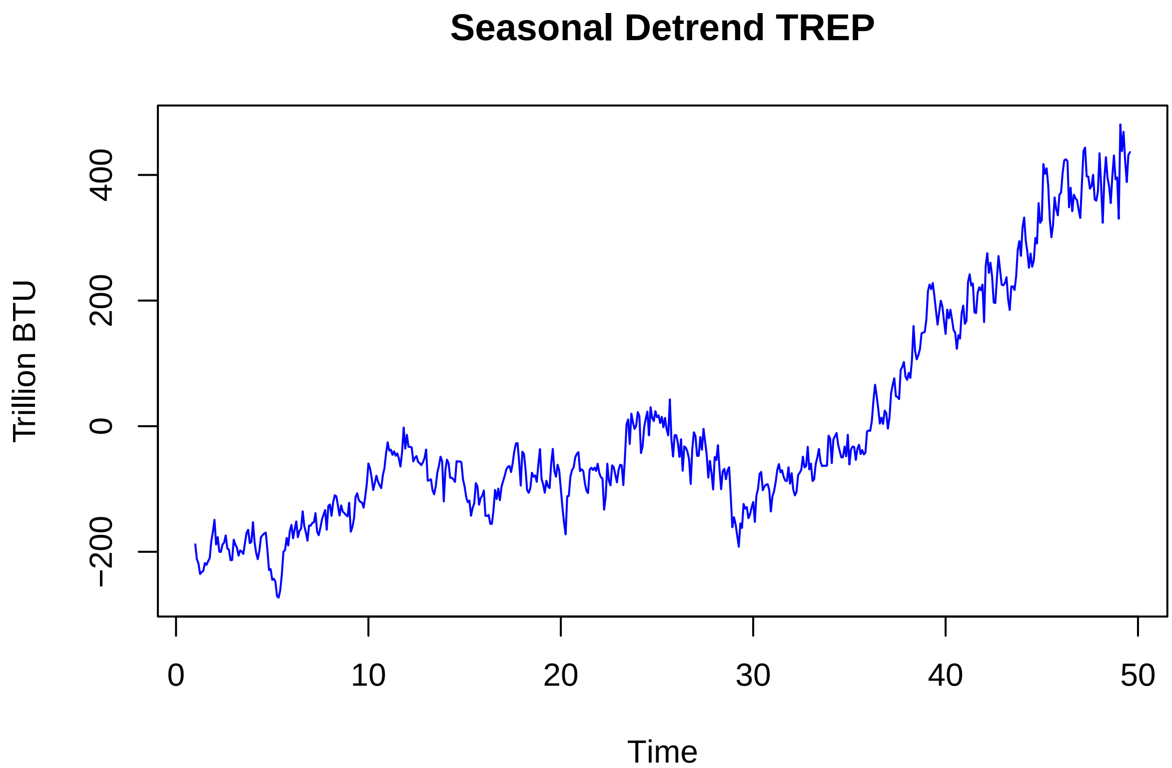
Q7

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? The only difference I was able to see was that there seemed to be less variabilty, but it did not seem to have that much of a difference. For HPC, there appeared to be more of a difference as there is more variability.

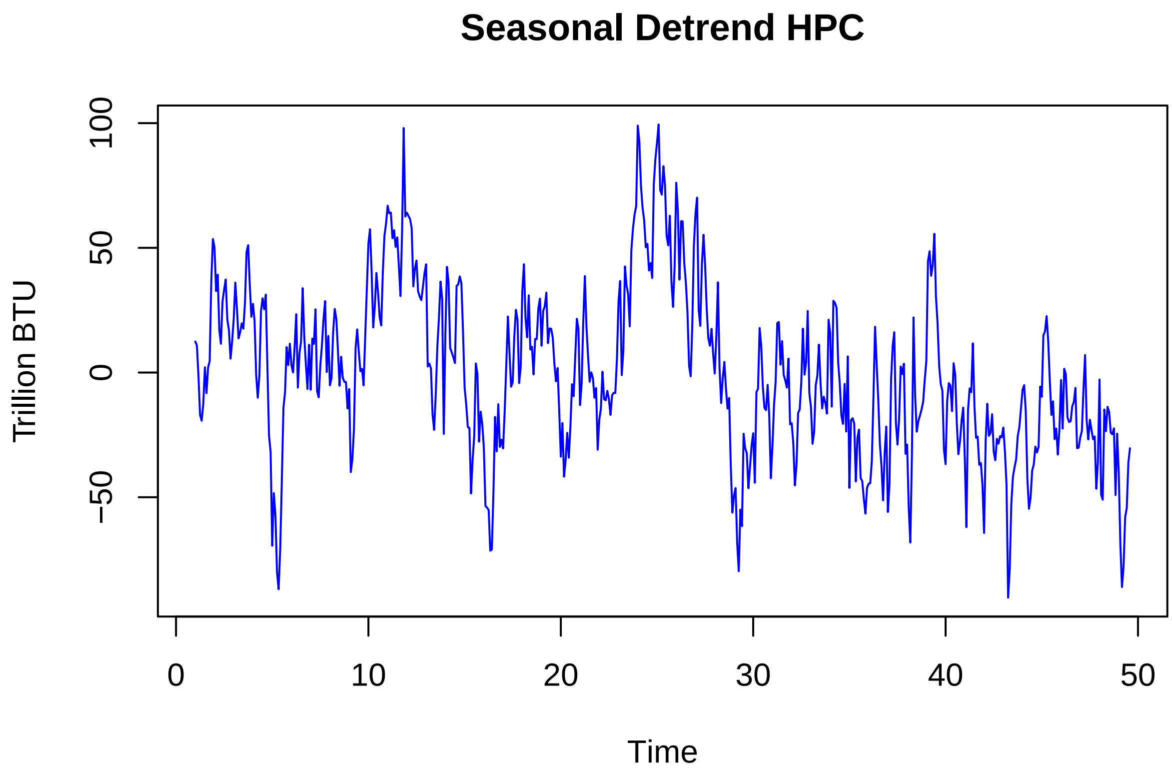
```r
TBEP_seas_comp=array(0,nobs)
for(i in 1:nobs){
  TBEP_seas_comp[i]=(beta_int_TBEP+beta_coeff_TBEP%*%dummies[i,])
}
deseason_energy_data_TBEP <- ts_energy_data[,1]-TBEP_seas_comp

TREP_seas_comp=array(0,nobs)
for(i in 1:nobs){
  TREP_seas_comp[i]=(beta_int_TREP+beta_coeff_TREP%*%dummies[i,])
}
deseason_energy_data_TREP <- ts_energy_data[,2]-TREP_seas_comp

HPC_seas_comp=array(0,nobs)
for(i in 1:nobs){
  HPC_seas_comp[i]=(beta_int_HPC+beta_coeff_HPC%*%dummies[i,])
}
deseason_energy_data_HPC <- ts_energy_data[,3]-HPC_seas_comp

plot(deseason_energy_data_TBEP,col="blue",ylab="Trillion BTU",main="Seasonal Detrend TBEP")
```




```r
plot(deseason_energy_data_TREP,col="blue",ylab="Trillion BTU",main="Seasonal Detrend TREP")
```



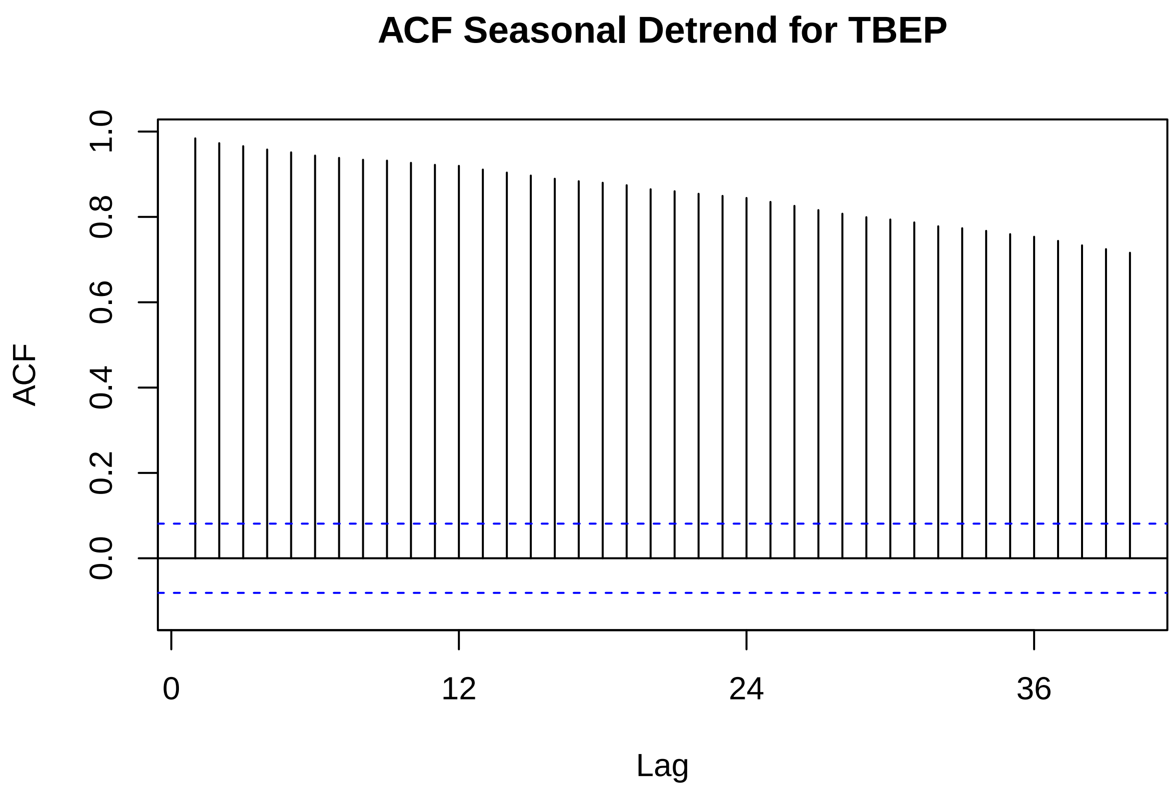

```
plot(deseason_energy_data_HPC,col="blue",ylab="Trillion BTU",main="Seasonal Detrend HPC")
```



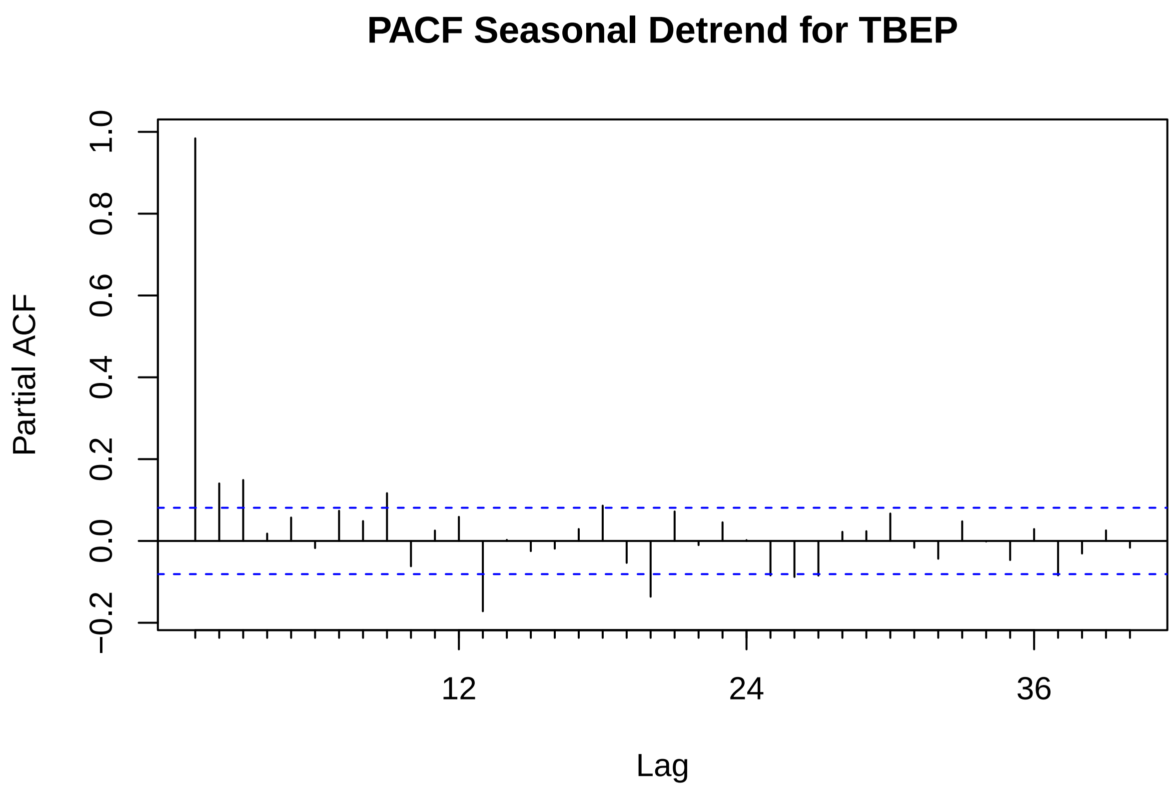
Q8

Plot ACF and PACF for the deseason series and compare with the plots from Q1. Did the plots change? How? HPC's ACF plot seemed to change the most with most of the seasonability been removed and being replaced with a straight downward trend. Both TBEP and TBEP did not have that much of a difference for their ACF graphs, which may be due to their seasonability regression output being smaller than for TBEP. The PCF graphs for all three seemed to have less variability.

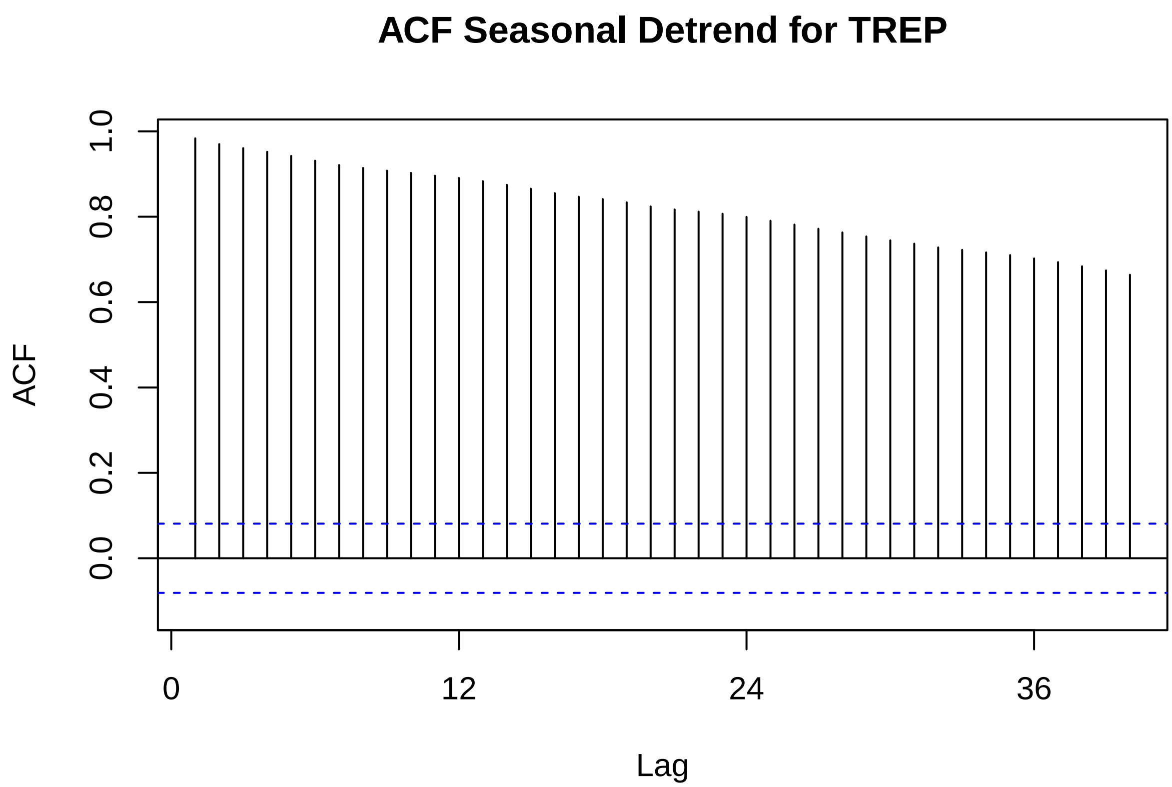
```
Acf(deseason_energy_data_TBEP,lag.max=40,main="ACF Seasonal Detrend for TBEP")
```



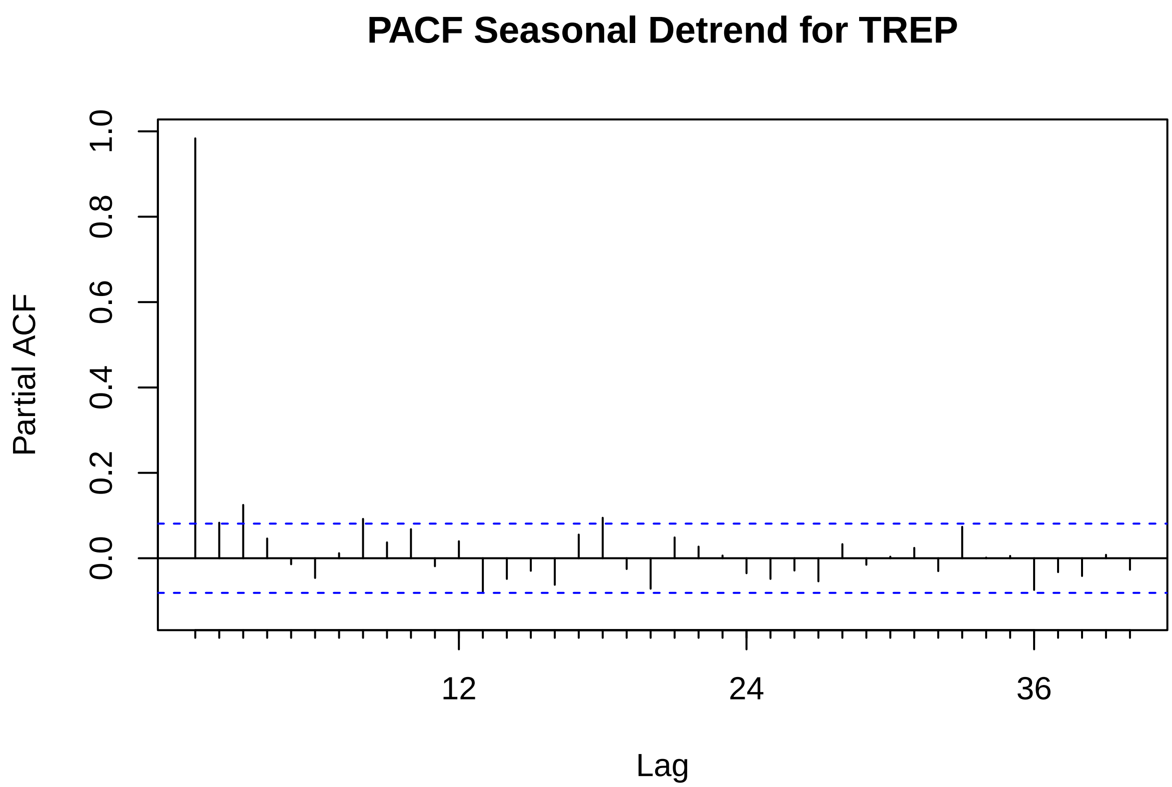

```r
Pacf(deseason_energy_data_TBEP,lag.max=40,main="PACF Seasonal Detrend for TBEP")
```



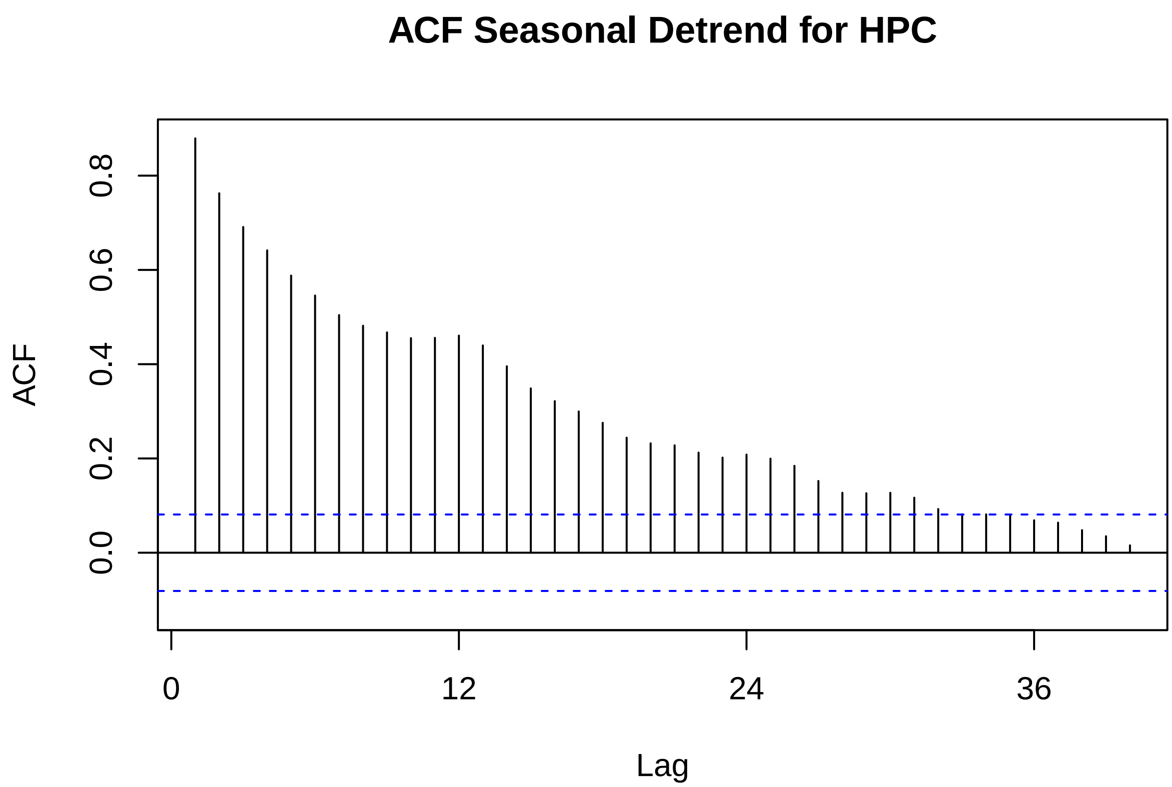

```
Acf(deseason_energy_data_TREP,lag.max=40,main="ACF Seasonal Detrend for TREP")
```



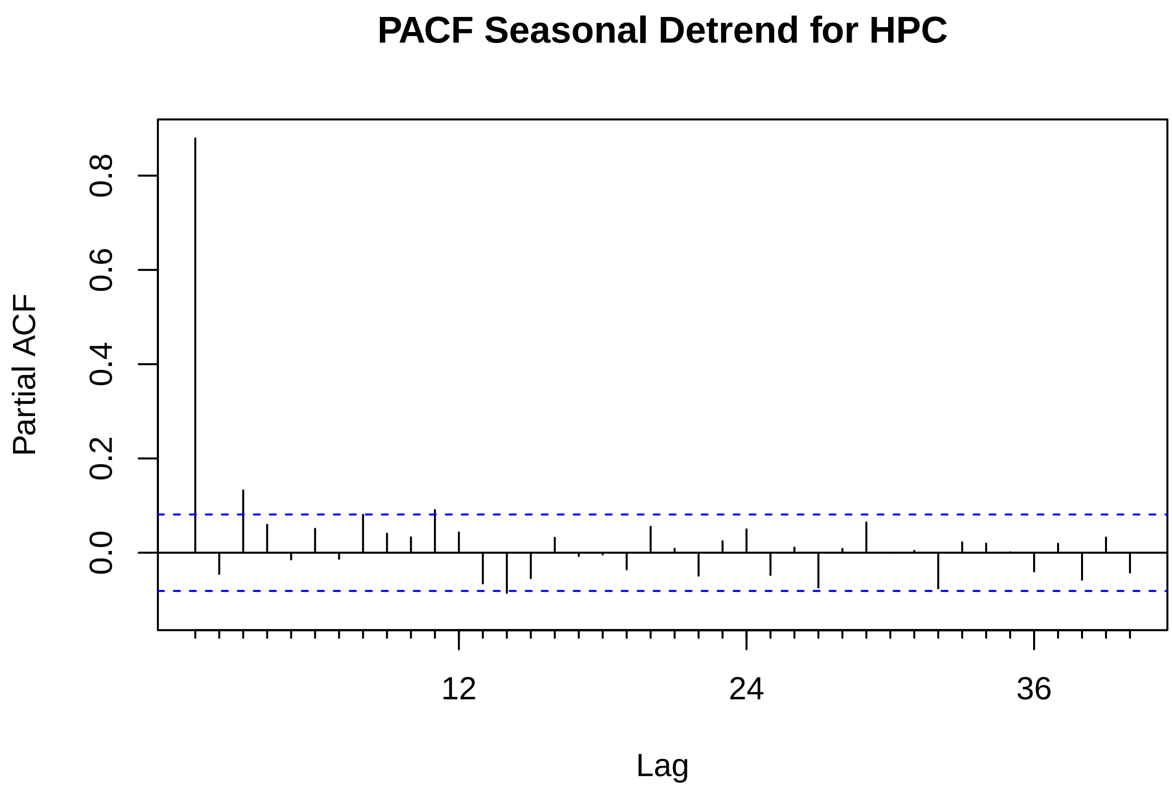

```
Pacf(deseason_energy_data_TREP,lag.max=40,main="PACF Seasonal Detrend for TREP")
```




```
Acf(deseason_energy_data_HPC,lag.max=40,main="ACF Seasonal Detrend for HPC")
```




```
Pacf(deseason_energy_data_HPC,lag.max=40,main="PACF Seasonal Detrend for HPC")
```


PACF Seasonal Detrend for HPC
Partial ACF
0.0
0.2
0.4
0.6
0.8
12
24
36
Lag