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

Assignment 3 - Due date 02/01/24

Yilun Zhu

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_Consumpti 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(lubridate)
```

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

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

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

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

```r
library(Kendall)
library(tseries)

#Import Data
Renew_Eng <- read.csv(file="../Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source.cs
#Take three column
V_Renew_Eng <- Renew_Eng[,4:6]
#Transfer Ren_Date from a factor to date, in order to use in plot
Ren_Date <- ym(Renew_Eng[,1])
Renew_Eng <- cbind(Ren_Date, V_Renew_Eng)
nobs <- nrow(Renew_Eng)
ncolu <- ncol(Renew_Eng)-1
Ren_Date <- Renew_Eng[,1]
ts_Renew_Eng <- ts(Renew_Eng[,2:4], start = c(1973,1), frequency = 12)
```

##Trend Component

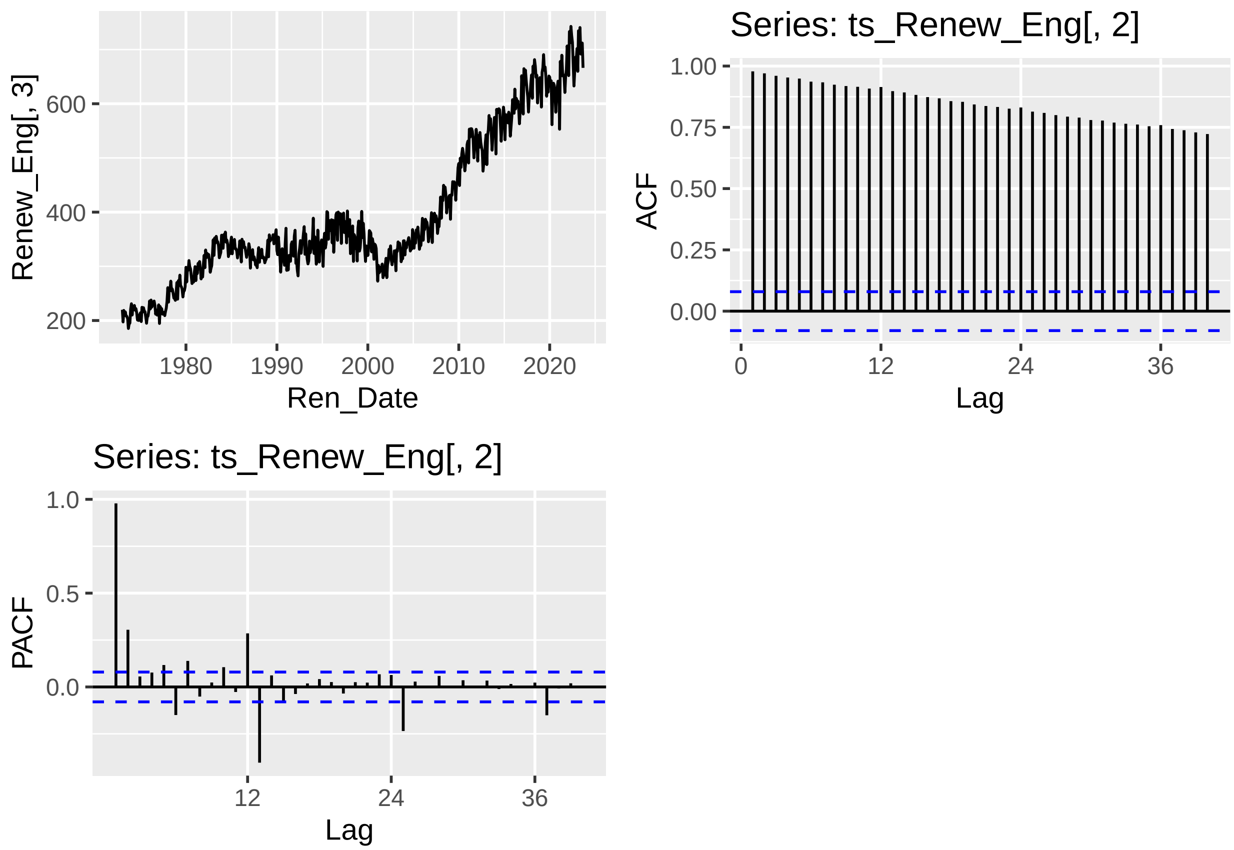
Q1

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)

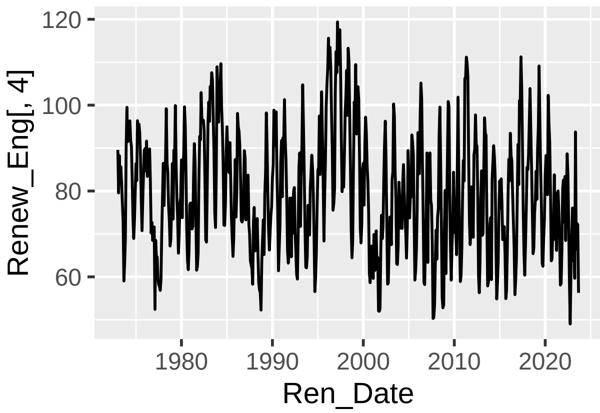
```r
library(cowplot)
```

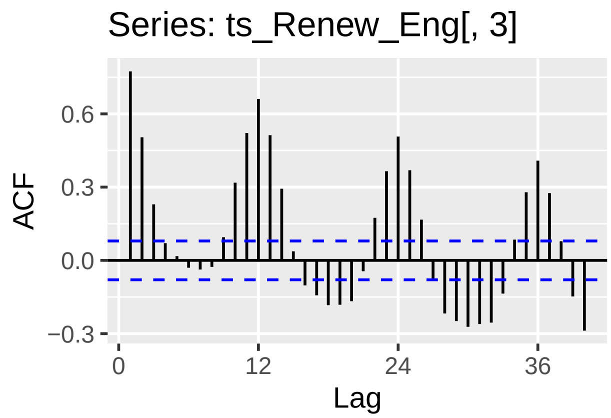
```
##
## Attaching package: 'cowplot'

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

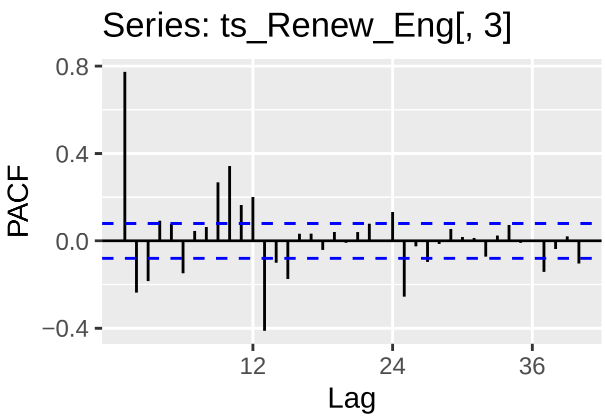
```r
TREPacf <-  Acf(ts_Renew_Eng[,2],lag.max=40,main=paste("TREPAcf"), plot = FALSE)
TREPpacf <- Pacf(ts_Renew_Eng[,2],lag.max=40,main=paste("TREPPacf"), plot = FALSE)
HPCacf <- Acf(ts_Renew_Eng[,3],lag.max=40,main=paste("HPCAcf"), plot = FALSE)
HPCpacf <- Pacf(ts_Renew_Eng[,3],lag.max=40,main=paste("HPCPacf"), plot = FALSE)

tsTREP_plot <- ggplot(Renew_Eng, aes(x=Ren_Date))+
  geom_line(aes(y=Renew_Eng[,3]),color = "black")
tsHPC_plot <- ggplot(Renew_Eng, aes(x=Ren_Date))+
  geom_line(aes(y=Renew_Eng[,4]),color = "black")

plot_grid(tsTREP_plot,autoplot(TREPacf),autoplot(TREPpacf))
```




```
plot_grid(tsHPC_plot,autoplot(HPCacf),autoplot(HPCpacf))
```







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?

```
#Total Biomass Energy Production and Total Renewable Energy Production all have a declining trend. Hydr
```

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.

```
t <- 1:nobs

iTREP <- 3
linear_Trend_TREP <- lm(Renew_Eng[,iTREP]~t)
summary(linear_Trend_TREP)
```

```
##
## Call:
## lm(formula = Renew_Eng[, iTREP] ~ 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
```

```
TREP_beta0 <- linear_Trend_TREP$coefficients[1]
TREP_beta1 <- linear_Trend_TREP$coefficients[2]
iHPC <- 4
linear_Trend_HPC <- lm(Renew_Eng[,iHPC]~t)
summary(linear_Trend_HPC)
```

```
##
## Call:
## lm(formula = Renew_Eng[, iHPC] ~ 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
```

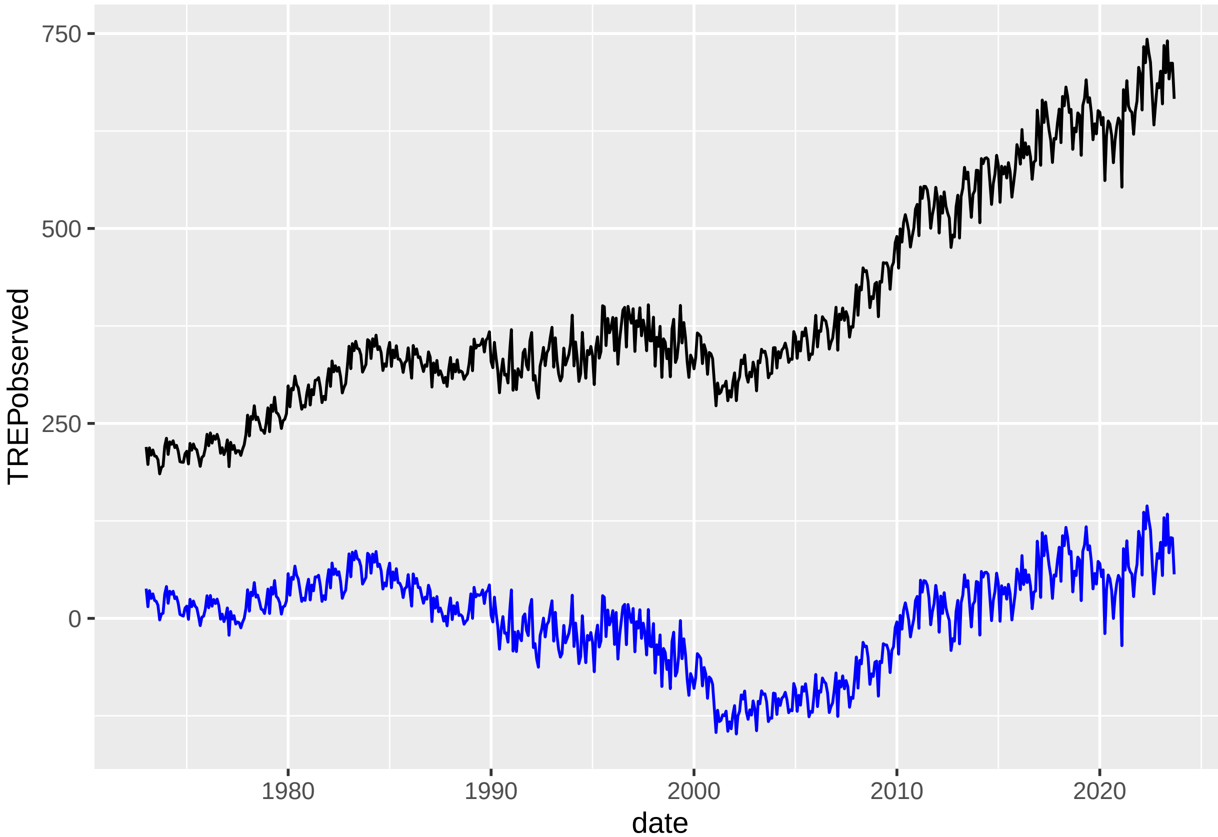
```
HPC_beta0 <- linear_Trend_HPC$coefficients[1]
HPC_beta1 <- linear_Trend_HPC$coefficients[2]

#Interpret: the beta1 for each linear model is the slope, which means each unit increase in independent
```

Q4

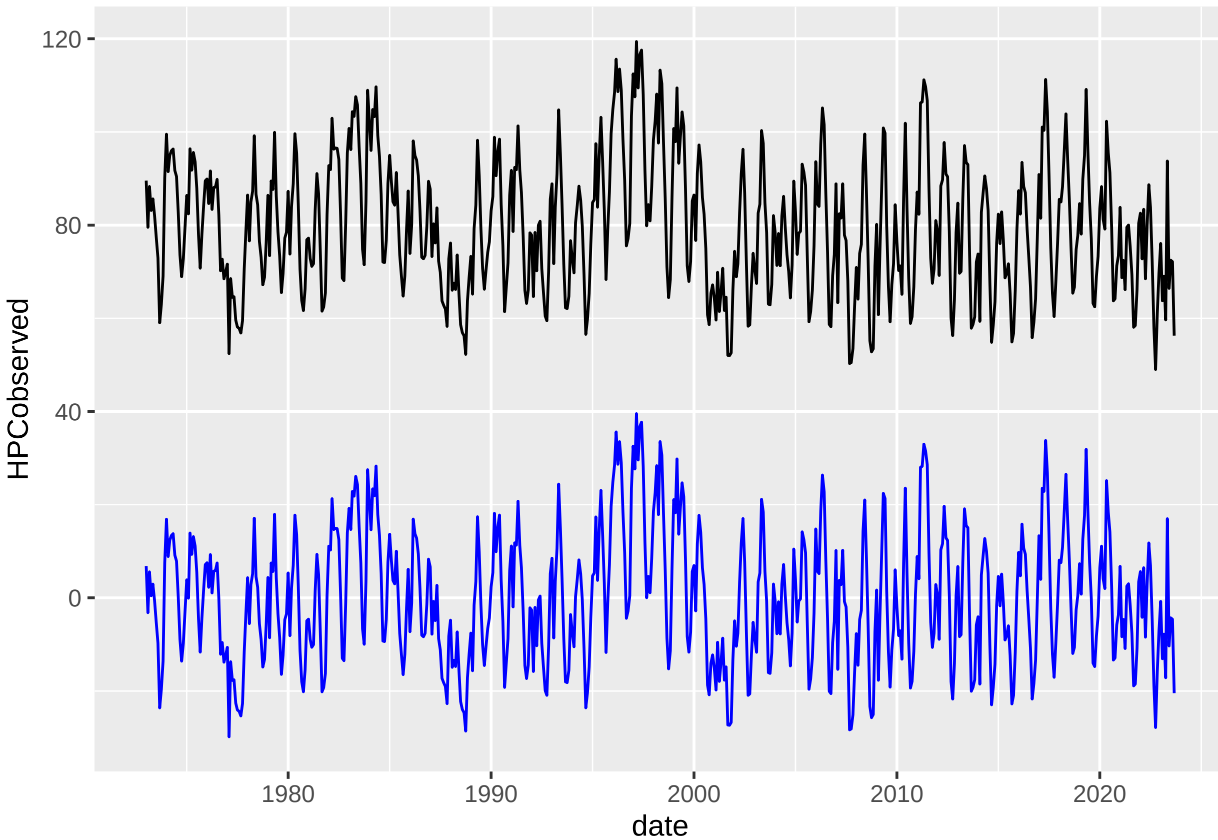
Use the regression coefficients from Q3 to detrend the series. Plot the detrended series and compare with the plots from Q1. What happened? Did anything change?

```
TREP_detrend <- Renew_Eng[,iTREP] - (TREP_beta0 + TREP_beta1*t)
df_TREPdetrend <- data.frame("date"= Renew_Eng$Ren_Date, "TREPobserved" = Renew_Eng[,iTREP], "TREPdetre
ggplot(df_TREPdetrend, aes(x=date))+
```

```r
  geom_line(aes(y=TREPobserved),color = "black")+
  geom_line(aes(y=TREPdetrend), color = "blue")
```




```r
HPC_detrend <- Renew_Eng[,iHPC] - (HPC_beta0 + HPC_beta1*t)
df_HPCdetrend <- data.frame("date"= Renew_Eng$Ren_Date, "HPCobserved" = Renew_Eng[,iHPC], "HPCdetrend"
ggplot(df_HPCdetrend, aes(x=date))+
  geom_line(aes(y=HPCobserved),color = "black")+
  geom_line(aes(y=HPCdetrend), color = "blue")
```

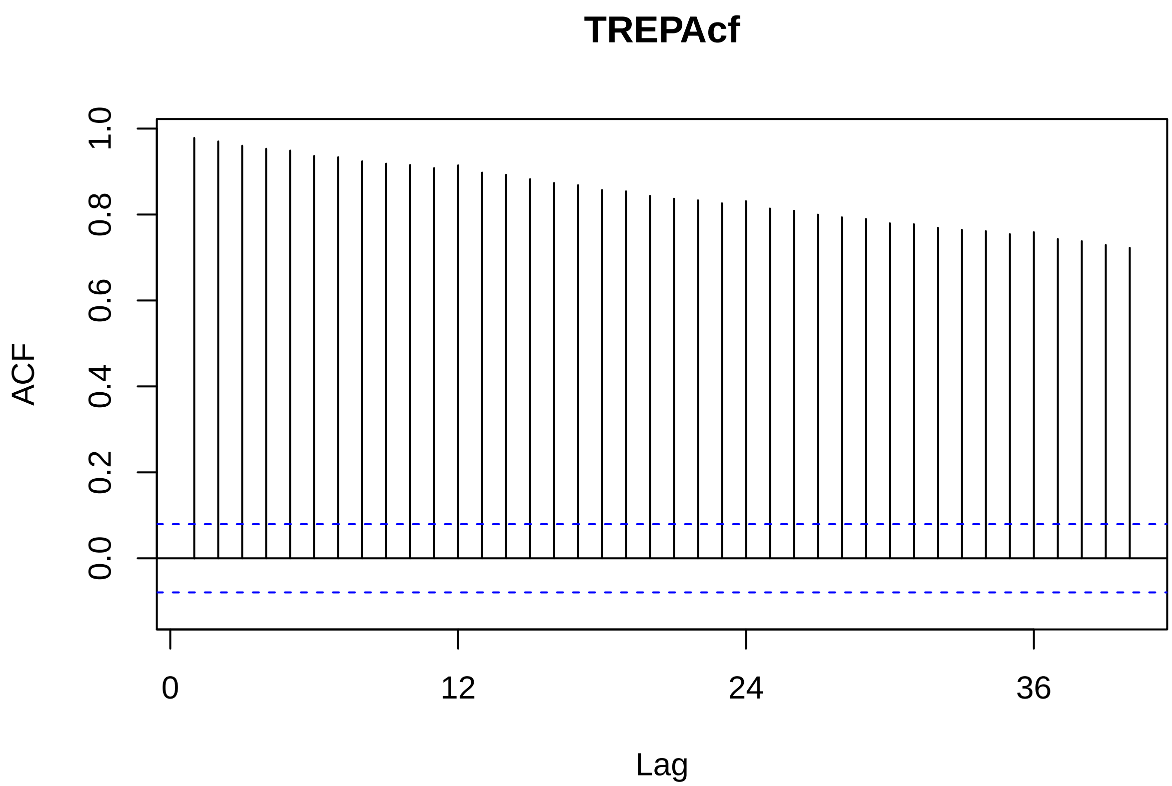



```
#In HPC plot, all the change is the line moves downward in some extent. In TREP plot, the line also mov
```

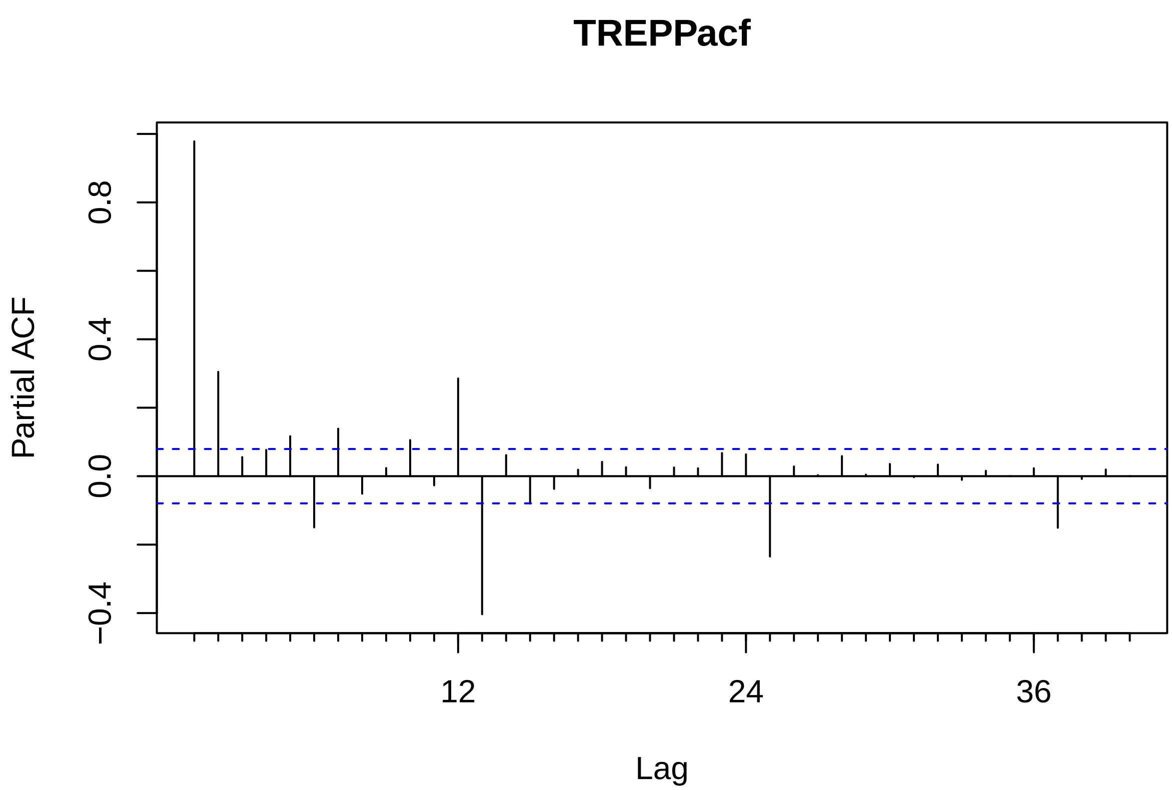
Q5

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?

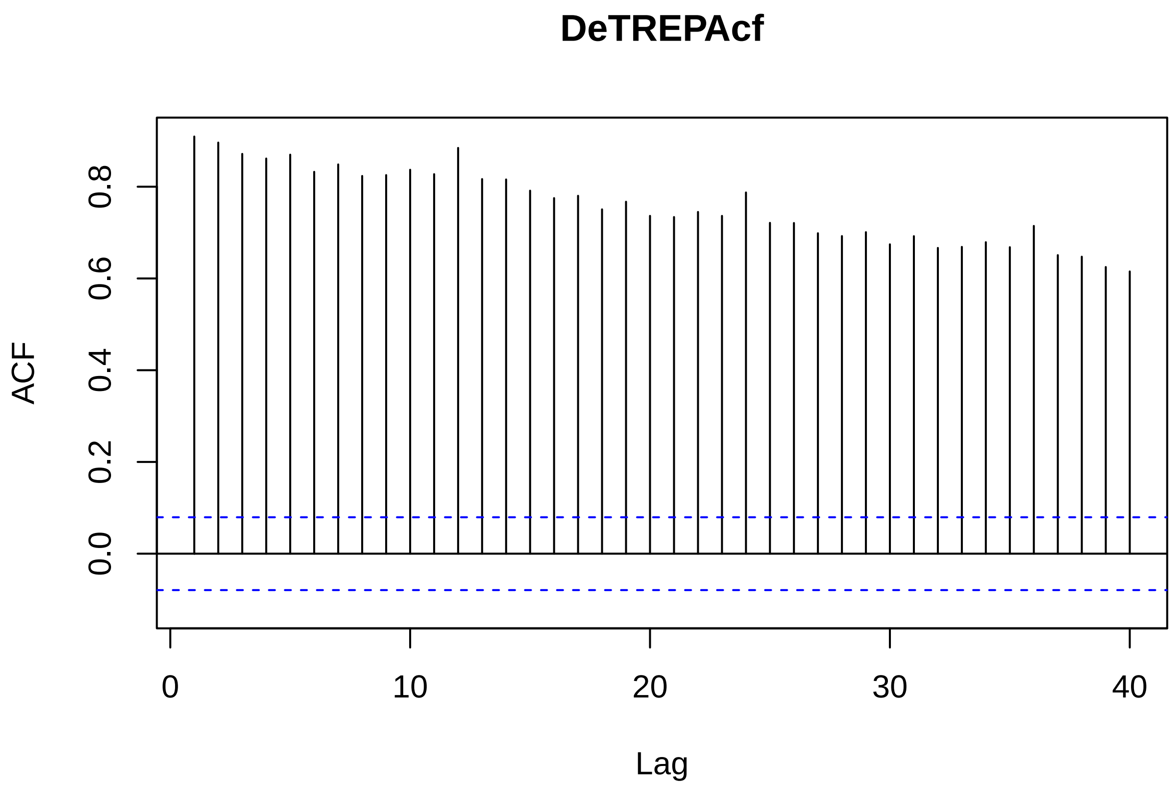
```
TREPacf <- Acf(ts_Renew_Eng[,2],lag.max=40,main=paste("TREP","Acf",sep=""))
```



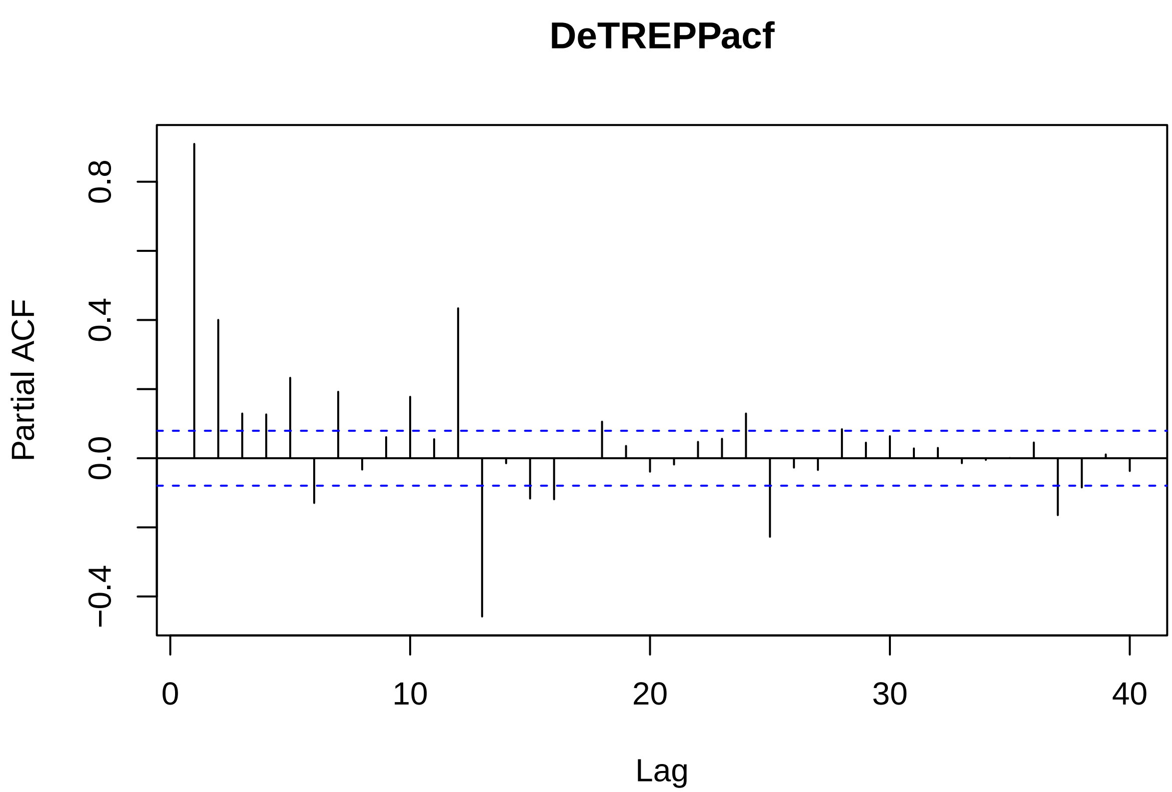

```
TREPpacf <- Pacf(ts_Renew_Eng[,2],lag.max=40,main=paste("TREP","Pacf",sep=""))
```



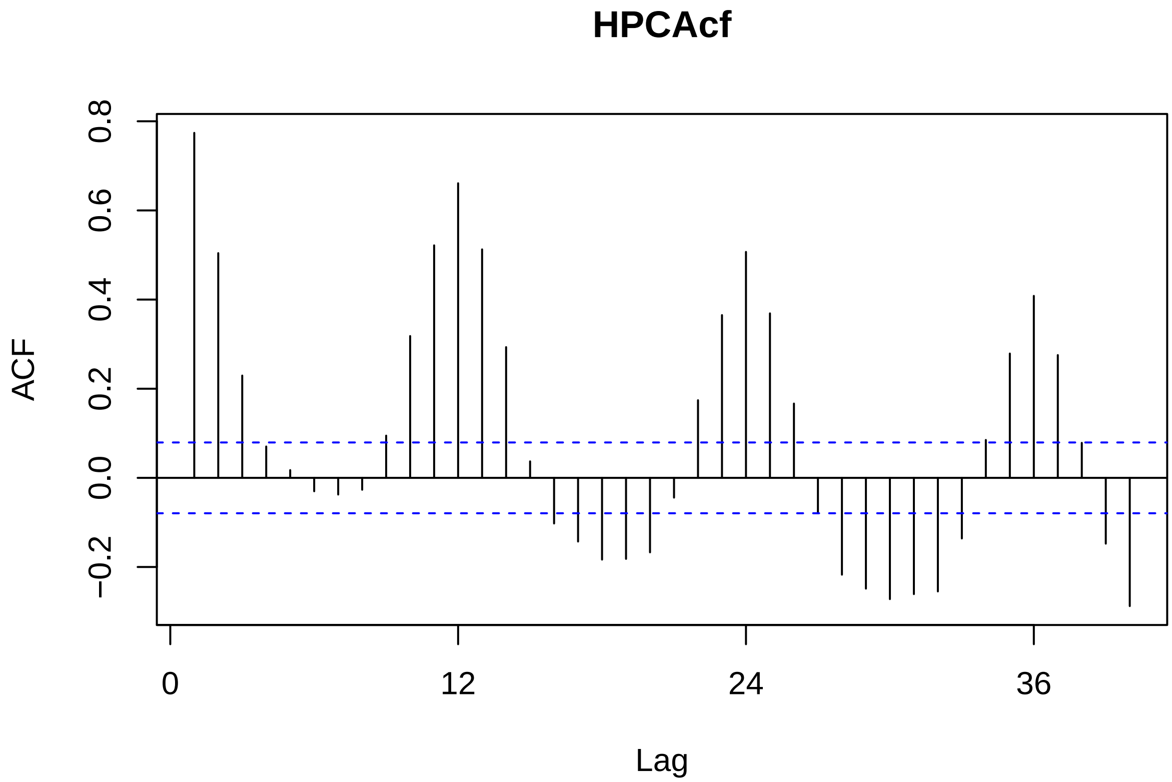
```r
DeTREPacf <- Acf(df_TREPdetrend[,3],lag.max=40,main=paste("DeTREP","Acf",sep=""))
```



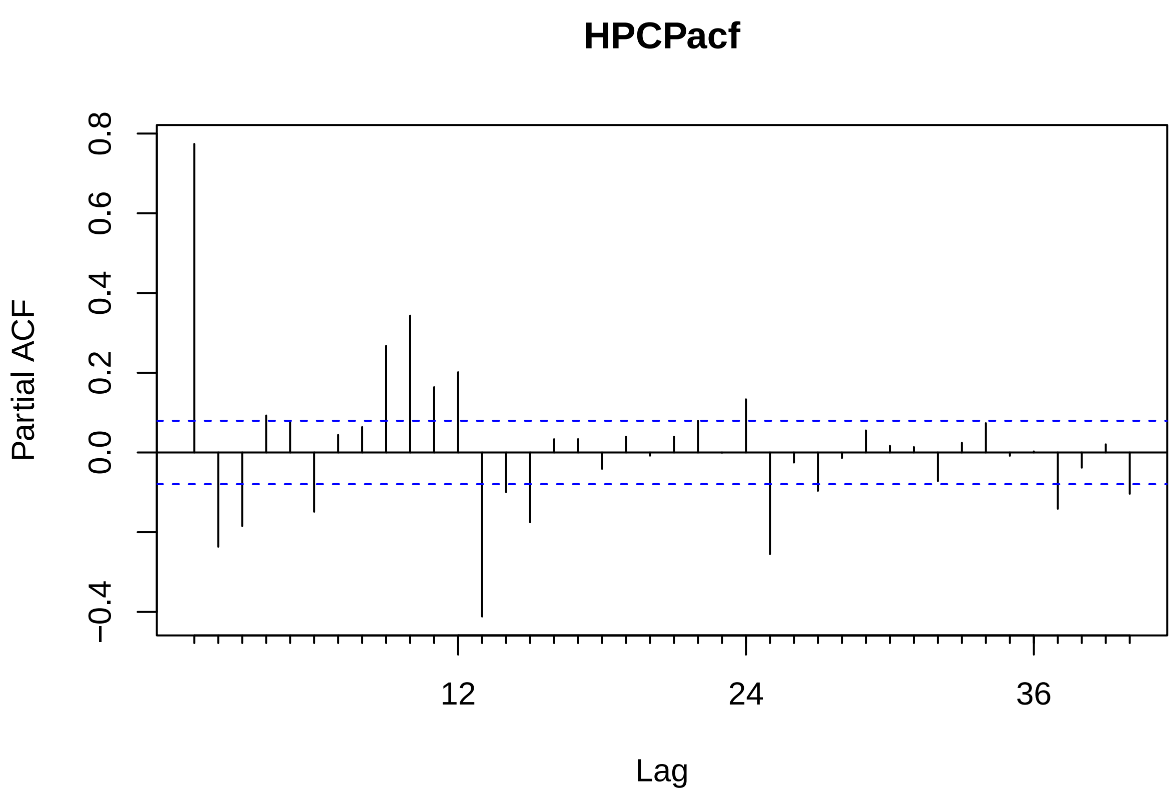

```r
DeTREPpacf <- Pacf(df_TREPdetrend[,3],lag.max=40,main=paste("DeTREP","Pacf",sep=""))
```



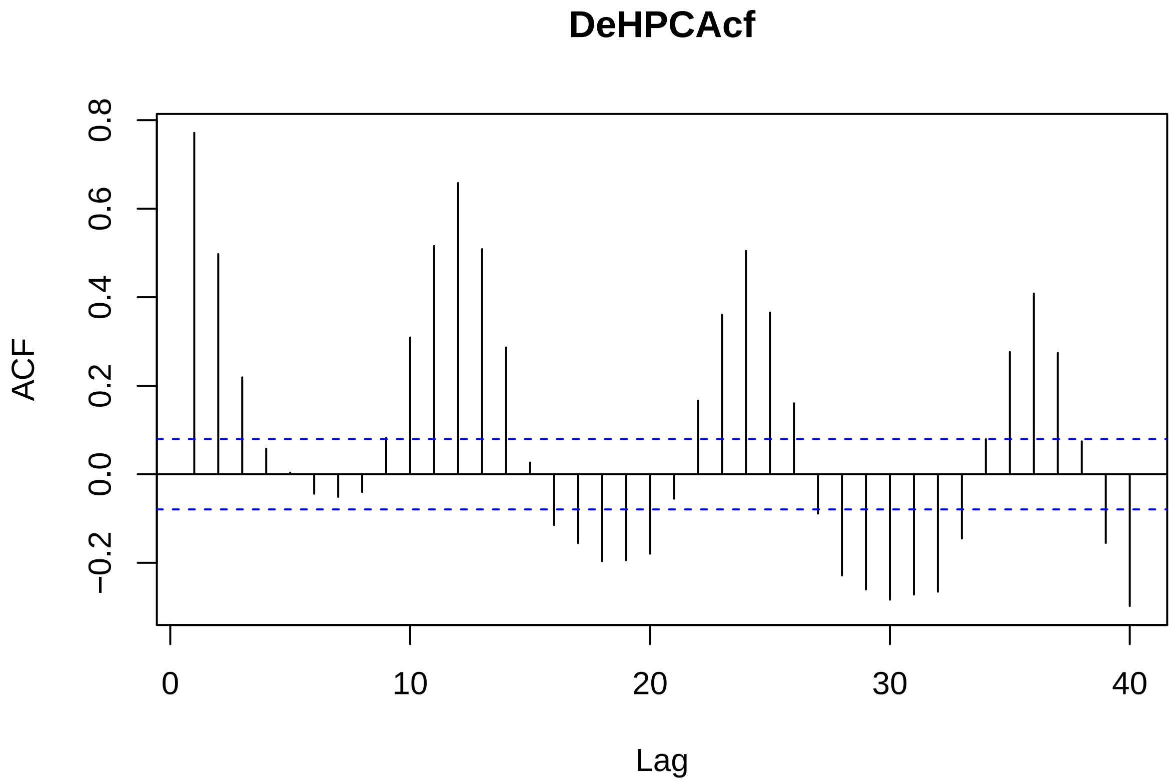
```r
HPCacf <- Acf(ts_Renew_Eng[,3],lag.max=40,main=paste("HPC","Acf",sep=""))
```



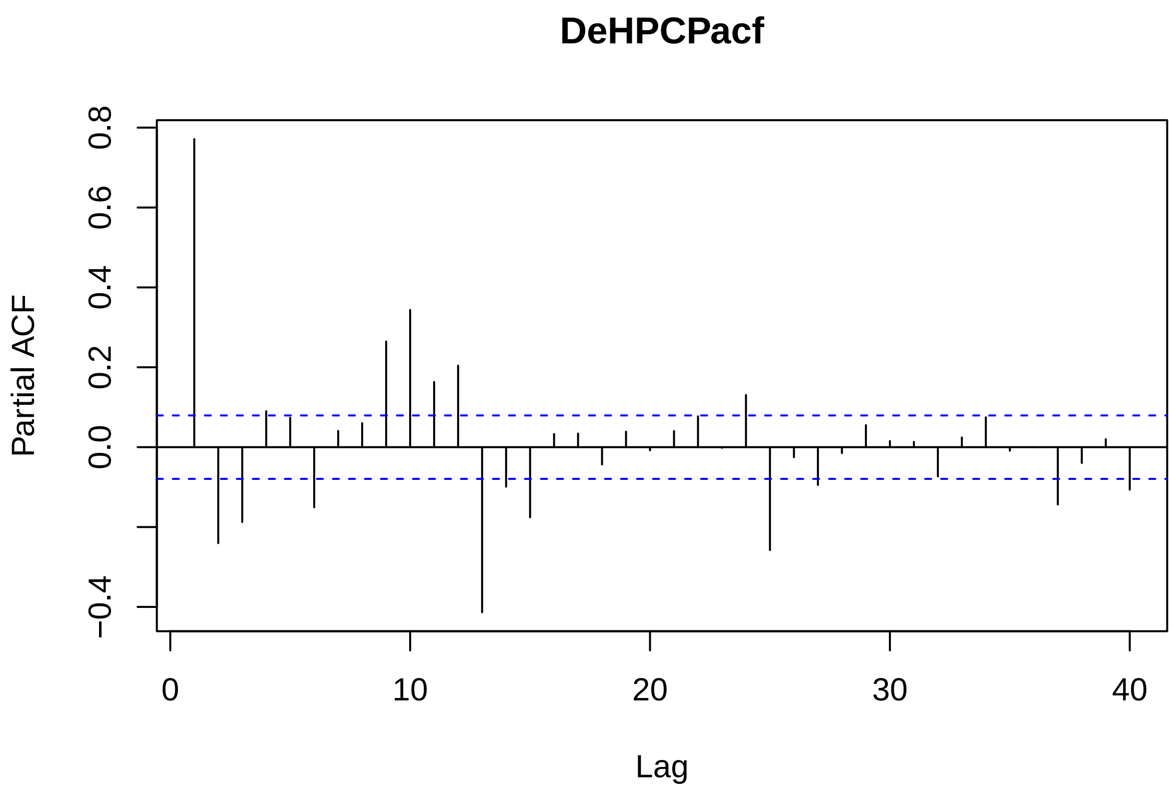

```r
HPCpacf <- Pacf(ts_Renew_Eng[,3],lag.max=40,main=paste("HPC","Pacf",sep=""))
```



```r
DeHPCacf <- Acf(df_HPCdetrend[,3],lag.max=40,main=paste("DeHPC","Acf",sep=""))
```




```r
DeHPCpacf <- Pacf(df_HPCdetrend[,3],lag.max=40,main=paste("DeHPC","Pacf",sep=""))
```



```
# These plots change but all change slightly. For TREP, more higher acf and pacf shows after detrend. F
```

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.

```
#For HPC, there is an obvious seasonal trend. For TREP, there seems to have a small decreasing seasonal
```

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?

```
dummies <- seasonaldummy(ts_Renew_Eng[,2])
TREPseas_linear_model <- lm(Renew_Eng[,3]~dummies)
summary(TREPseas_linear_model)
```

```
##
## Call:
## lm(formula = Renew_Eng[, 3] ~ dummies)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -199.19  -86.35  -48.84  113.18  331.58
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  404.526     19.574  20.666   <2e-16 ***
## dummiesJan     2.962     27.546   0.108    0.914
## dummiesFeb   -34.476     27.546  -1.252    0.211
## dummiesMar     3.929     27.546   0.143    0.887
## dummiesApr    -8.695     27.546  -0.316    0.752
## dummiesMay     6.645     27.546   0.241    0.809
## dummiesJun    -4.198     27.546  -0.152    0.879
## dummiesJul     2.460     27.546   0.089    0.929
## dummiesAug    -5.026     27.546  -0.182    0.855
## dummiesSep   -29.119     27.546  -1.057    0.291
## dummiesOct   -20.068     27.682  -0.725    0.469
## dummiesNov   -20.346     27.682  -0.735    0.463
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 138.4 on 597 degrees of freedom
## Multiple R-squared:  0.009296,   Adjusted R-squared:  -0.008958
## F-statistic: 0.5093 on 11 and 597 DF,  p-value: 0.8976
```

```r
HPCseas_linear_model <- lm(Renew_Eng[,4]~dummies)
summary(HPCseas_linear_model)
```

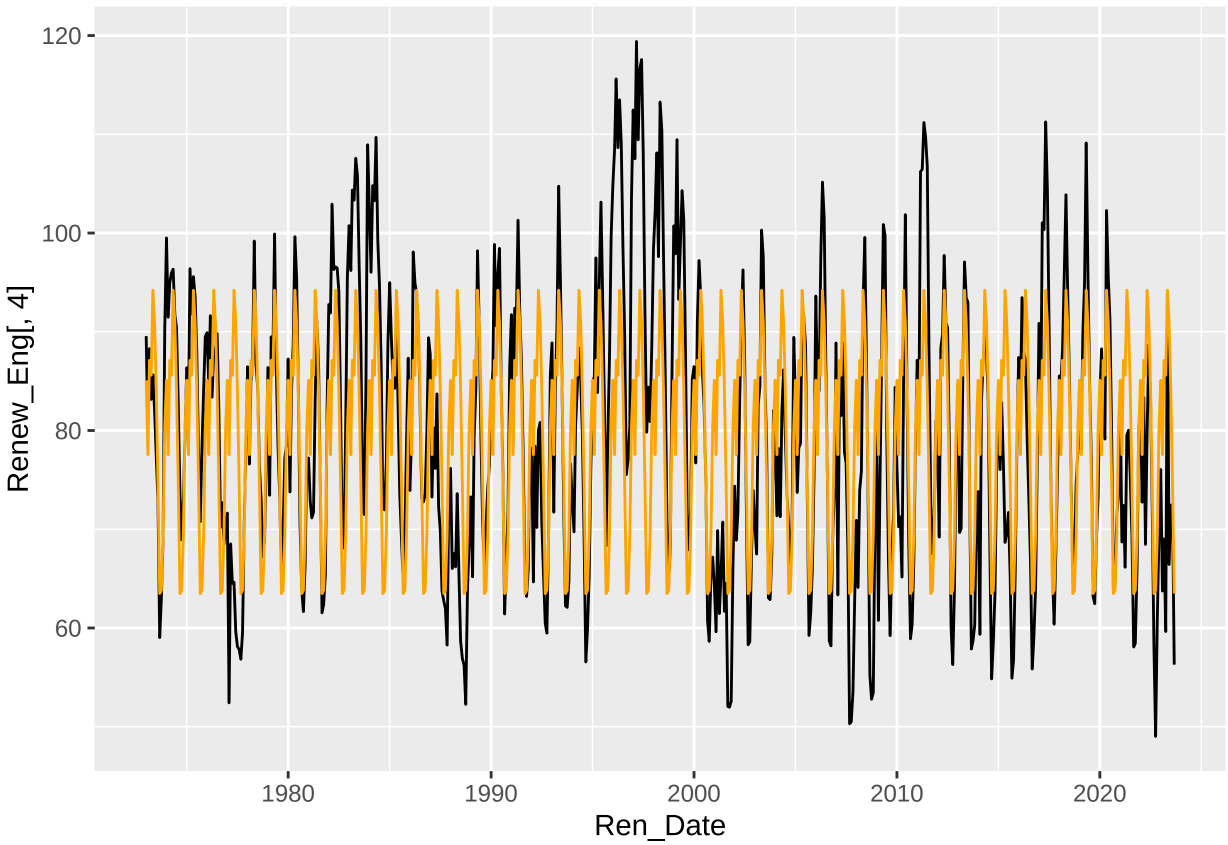
```
##
## Call:
## lm(formula = Renew_Eng[, 4] ~ dummies)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -31.323  -5.849  -0.468   6.243  32.290
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   80.282      1.470  54.601  < 2e-16 ***
## dummiesJan     4.807      2.069   2.323  0.02050 *
## dummiesFeb    -2.725      2.069  -1.317  0.18831
## dummiesMar     6.825      2.069   3.298  0.00103 **
## dummiesApr     5.319      2.069   2.571  0.01039 *
## dummiesMay    13.922      2.069   6.729 4.02e-11 ***
## dummiesJun    10.650      2.069   5.147 3.60e-07 ***
## dummiesJul     3.912      2.069   1.891  0.05914 .
## dummiesAug    -5.677      2.069  -2.744  0.00626 **
## dummiesSep   -16.797      2.069  -8.118 2.72e-15 ***
## dummiesOct   -16.468      2.079  -7.920 1.17e-14 ***
## dummiesNov   -10.885      2.079  -5.235 2.29e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.4 on 597 degrees of freedom
## Multiple R-squared:  0.4697, Adjusted R-squared:  0.4599
## F-statistic: 48.07 on 11 and 597 DF,  p-value: < 2.2e-16
```

```r
#The result shows HPC has a seasonal trend and TREP doesn't since its regression has a great P value.
```

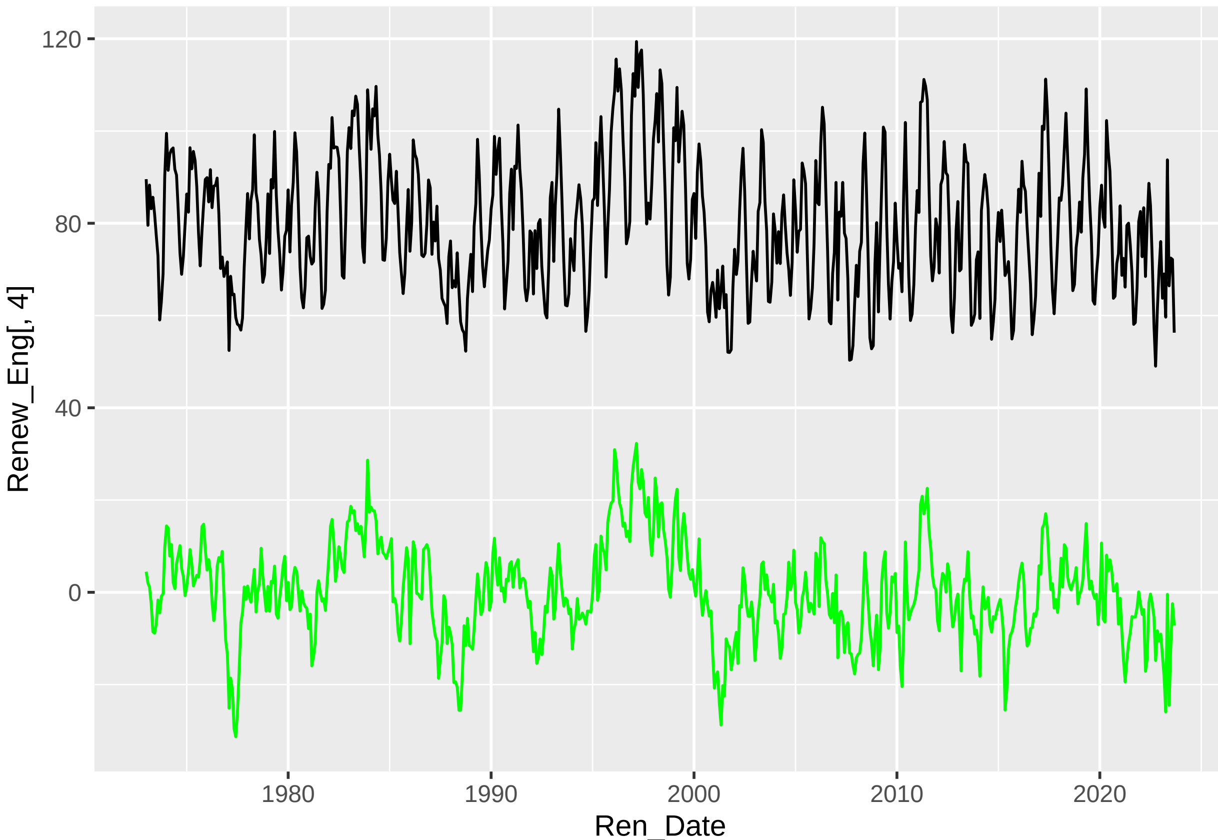
Q8

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?

```r
#Since TREP doesn't have the seasonal trend, so I just did deseason to HPC.
HPCbeta0 <- HPCseas_linear_model$coefficients[1]
HPCbeta1 <- HPCseas_linear_model$coefficients[2:12]

HPCseas_comp <- array(0,nobs)
for (i in 1:nobs) {
  HPCseas_comp[i] <- HPCbeta0 + HPCbeta1 %*% dummies[i,]
}

ggplot(Renew_Eng, aes(x=Ren_Date))+
  geom_line(aes(y=Renew_Eng[,4]), col = "black")+
  geom_line(aes(y= HPCseas_comp), col = "orange")
```




```r
HPC_deseason <- Renew_Eng[,4] - HPCseas_comp

ggplot(Renew_Eng, aes(x=Ren_Date))+
  geom_line(aes(y=Renew_Eng[,4]), col = "black")+
  geom_line(aes(y= HPC_deseason), col = "green")
```

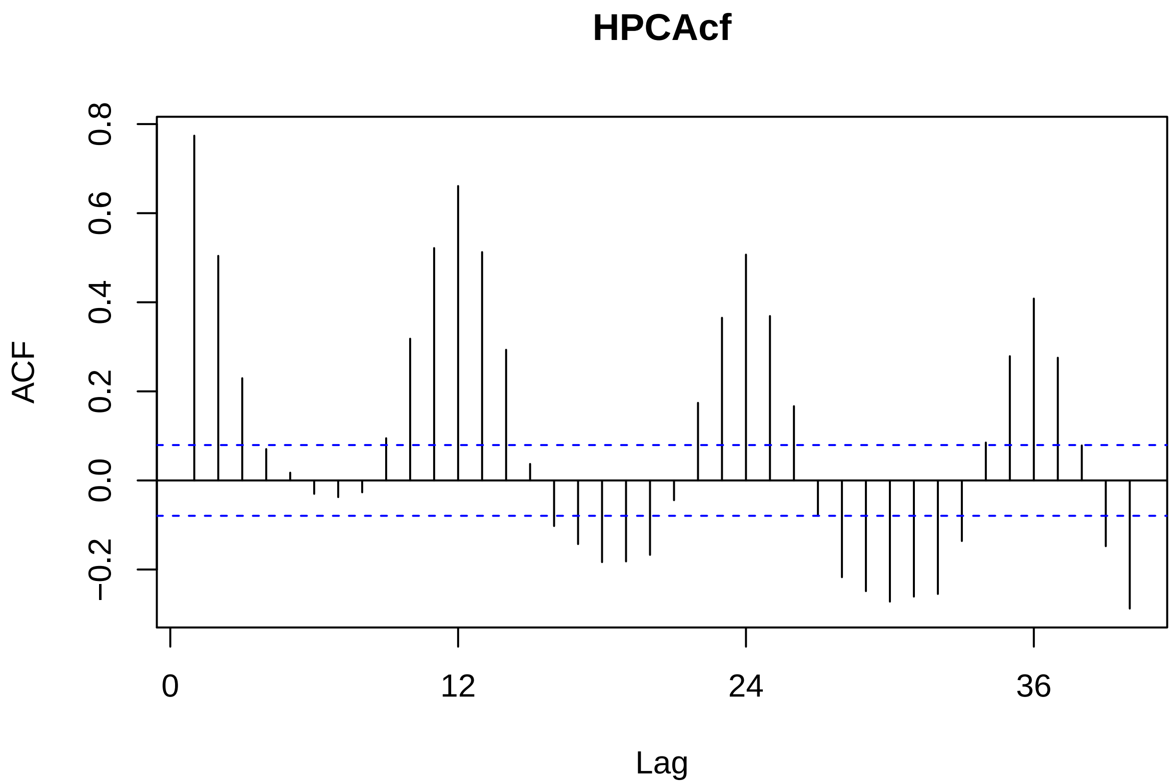



```
#The plots have moved downward as a whole, and the trend of HPC has changed in some time periods.
```

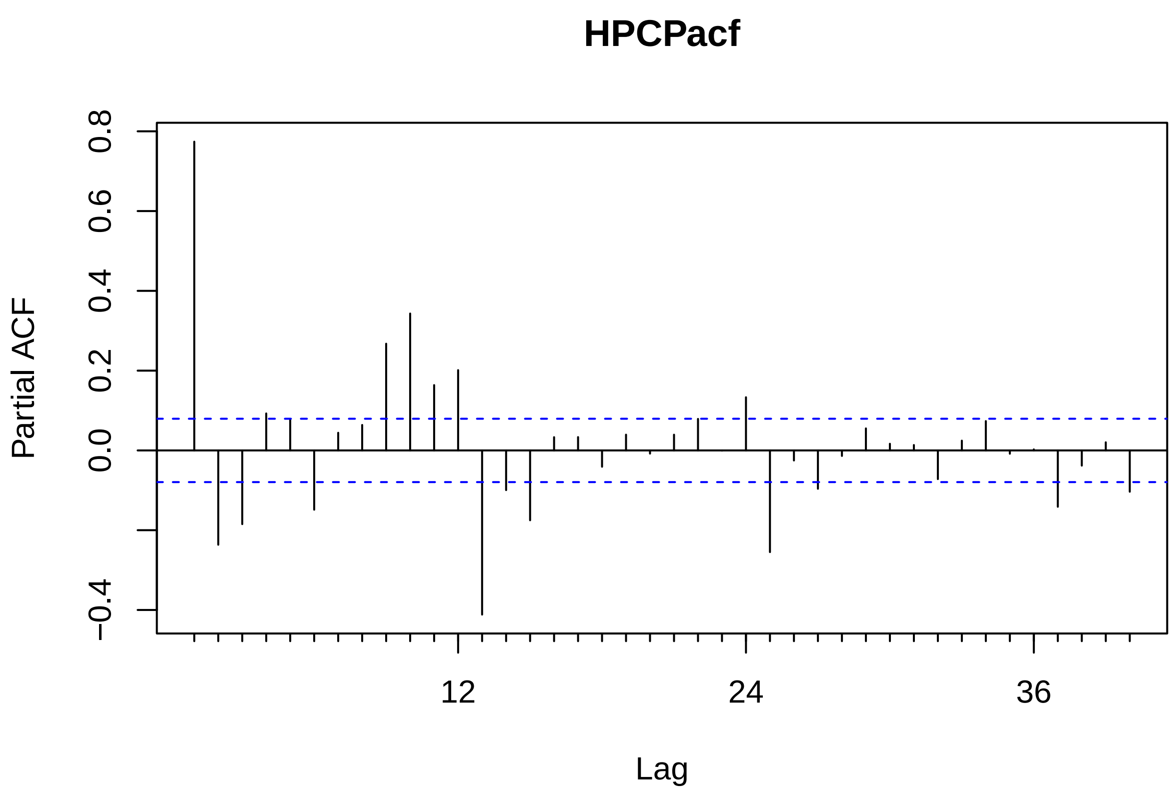
Q9

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?

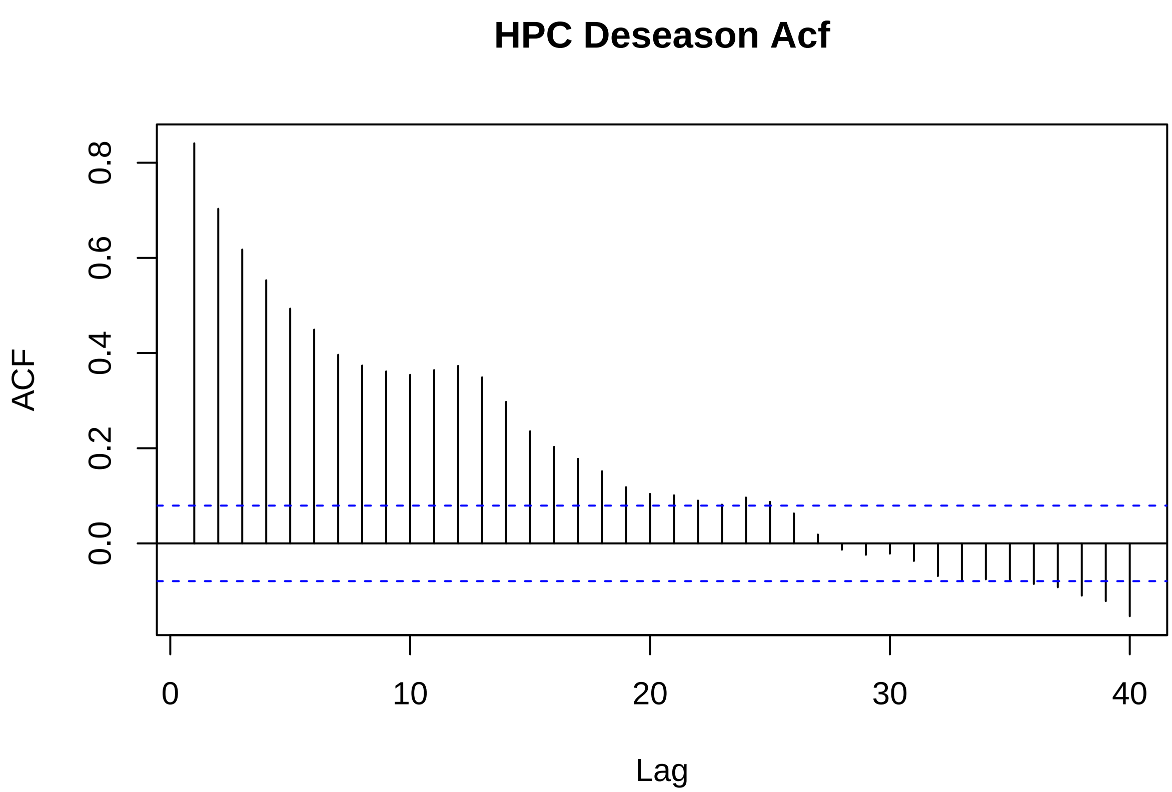
```
HPCacf <- Acf(ts_Renew_Eng[,3],lag.max=40,main=paste("HPC","Acf",sep=""))
```



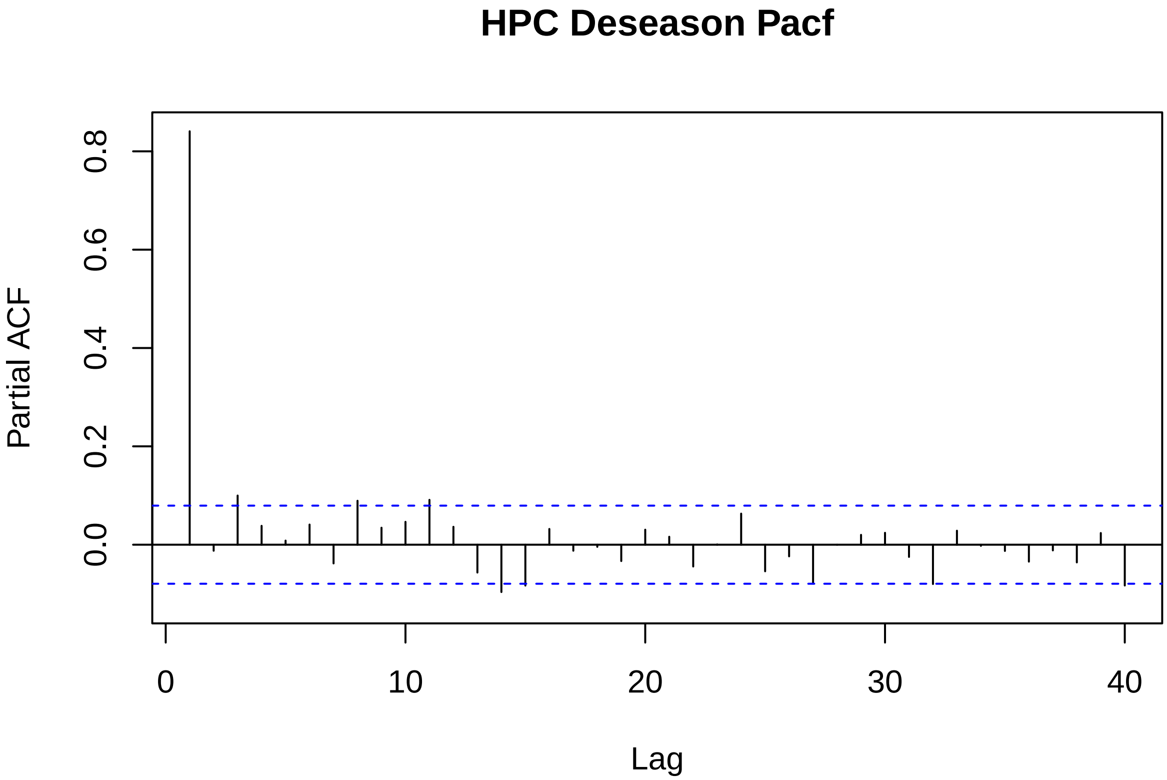

```
HPCpacf <- Pacf(ts_Renew_Eng[,3],lag.max=40,main=paste("HPC","Pacf",sep=""))
```



```
df_HPCseason <- data.frame("date"= Renew_Eng$Ren_Date, "HPCobserved" = Renew_Eng[,iHPC], "HPCseason" = R
HPCdes_acf <- Acf(df_HPCseason[,3],lag.max=40,main=paste("HPC","Deseason Acf",sep=" "))
```




```
HPCdes_pacf <- Pacf(df_HPCseason[,3],lag.max=40,main=paste("HPC","Deseason Pacf",sep=" "))
```




```
#Both Acf and Pacf change a lot. Deseasoned Acf no longer show seasonal change but a decreasing acf gra
```