

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

Assignment 3 - Due date 02/08/22

Yared S. Asfaw

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.

Loading packages

```
#Load/install required package here
library(lubridate)
library(ggplot2)
library(forecast)
library(Kendall)
library(tseries)
library(openxlsx)
```

```
#Importing the data set
library(openxlsx)
us_energy.data <- read.xlsx("./Data/Raw/Table_10.1_Ren._Energy_PC._by_Source.xlsx", startRow=11)

head(us_energy.data)
```

```
##      Month Wood.Energy.Production Biofuels.Production
## 1      NA      (Trillion Btu)      (Trillion Btu)
## 2 26665      129.63      Not Available
## 3 26696      117.194      Not Available
## 4 26724      129.763      Not Available
## 5 26755      125.462      Not Available
## 6 26785      129.624      Not Available
##      Total.Biomass.Energy.Production Total.Renewable.Energy.Production
## 1      (Trillion Btu)      (Trillion Btu)
## 2      129.787      403.981
## 3      117.338      360.9
## 4      129.938      400.161
## 5      125.636      380.47
## 6      129.834      392.141
##      Hydroelectric.Power.Consumption Geothermal.Energy.Consumption
## 1      (Trillion Btu)      (Trillion Btu)
## 2      272.703      1.491
## 3      242.199      1.363
## 4      268.81      1.412
## 5      253.185      1.649
## 6      260.77      1.537
##      Solar.Energy.Consumption Wind.Energy.Consumption Wood.Energy.Consumption
## 1      (Trillion Btu)      (Trillion Btu)      (Trillion Btu)
## 2      Not Available      Not Available      129.63
## 3      Not Available      Not Available      117.194
## 4      Not Available      Not Available      129.763
## 5      Not Available      Not Available      125.462
## 6      Not Available      Not Available      129.624
##      Waste.Energy.Consumption Biofuels.Consumption
## 1      (Trillion Btu)      (Trillion Btu)
## 2      0.157      Not Available
## 3      0.144      Not Available
## 4      0.176      Not Available
## 5      0.174      Not Available
## 6      0.21      Not Available
##      Total.Biomass.Energy.Consumption Total.Renewable.Energy.Consumption
## 1      (Trillion Btu)      (Trillion Btu)
## 2      129.787      403.981
## 3      117.338      360.9
## 4      129.938      400.161
## 5      125.636      380.47
## 6      129.834      392.141
```

```
# Select the three variables (Total Biomass Energy #Production (TEBP), Total Renewable Energy Production
us_energy.df <- data.frame(us_energy.data[-1,c(1,4:6)])
head(us_energy.df)
```

```
##      Month Total.Biomass.Energy.Production Total.Renewable.Energy.Production
## 2 26665                                129.787                        403.981
## 3 26696                                117.338                        360.9
## 4 26724                                129.938                        400.161
## 5 26755                                125.636                        380.47
## 6 26785                                129.834                        392.141
## 7 26816                                125.611                        377.232
##      Hydroelectric.Power.Consumption
## 2                                272.703
## 3                                242.199
## 4                                268.81
## 5                                253.185
## 6                                260.77
## 7                                249.859
```

```
colnames(us_energy.df)
```

```
## [1] "Month"                                "Total.Biomass.Energy.Production"
## [3] "Total.Renewable.Energy.Production" "Hydroelectric.Power.Consumption"
```

```
# renaming the column names for convenience
colnames(us_energy.df) <- c("my_date", "TBEP", "TREP", "HEPC")
colnames(us_energy.df)
```

```
## [1] "my_date" "TBEP"      "TREP"      "HEPC"
```

```
head(us_energy.df)
```

```
##      my_date      TBEP      TREP      HEPC
## 2 26665 129.787 403.981 272.703
## 3 26696 117.338 360.9 242.199
## 4 26724 129.938 400.161 268.81
## 5 26755 125.636 380.47 253.185
## 6 26785 129.834 392.141 260.77
## 7 26816 125.611 377.232 249.859
```

```
#Checking the class of the object created
is.data.frame(us_energy.df)
```

```
## [1] TRUE
```

```
# Checking the data types of the selected variables
sapply(us_energy.df, class)
```

```
##      my_date      TBEP      TREP      HEPC
##      "numeric" "character" "character" "character"
```

```
# formatting the my_date column to "Date"
```

```
start_date <- as.Date("01/01/1973", format = "%m/%d/%Y") # the date format in the dataset is m/d/y
us_energy.df$my_date <- seq(start_date, by= "month", length.out=585)
```

```
# Changing the data types into numeric
us_energy.df$TBEP <- as.numeric(us_energy.df$TBEP)
us_energy.df$TREP <- as.numeric(us_energy.df$TREP)
us_energy.df$HEPC <- as.numeric(us_energy.df$HEPC)
head(us_energy.df)
```

```
##      my_date    TBEP    TREP    HEPC
## 2 1973-01-01 129.787 403.981 272.703
## 3 1973-02-01 117.338 360.900 242.199
## 4 1973-03-01 129.938 400.161 268.810
## 5 1973-04-01 125.636 380.470 253.185
## 6 1973-05-01 129.834 392.141 260.770
## 7 1973-06-01 125.611 377.232 249.859
```

```
sapply(us_energy.df, class)
```

```
##      my_date      TBEP      TREP      HEPC
##      "Date" "numeric" "numeric" "numeric"
```

```
str(us_energy.df)
```

```
## 'data.frame': 585 obs. of 4 variables:
## $ my_date: Date, format: "1973-01-01" "1973-02-01" ...
## $ TBEP : num 130 117 130 126 130 ...
## $ TREP : num 404 361 400 380 392 ...
## $ HEPC : num 273 242 269 253 261 ...
```

```
tail(us_energy.df)
```

```
##      my_date    TBEP    TREP    HEPC
## 581 2021-04-01 393.384 1036.825 168.449
## 582 2021-05-01 425.888 1096.106 199.952
## 583 2021-06-01 414.266 1031.691 211.216
## 584 2021-07-01 429.335 986.802 194.009
## 585 2021-08-01 414.181 1003.261 183.850
## 586 2021-09-01 396.794 964.228 157.616
```

```
##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)

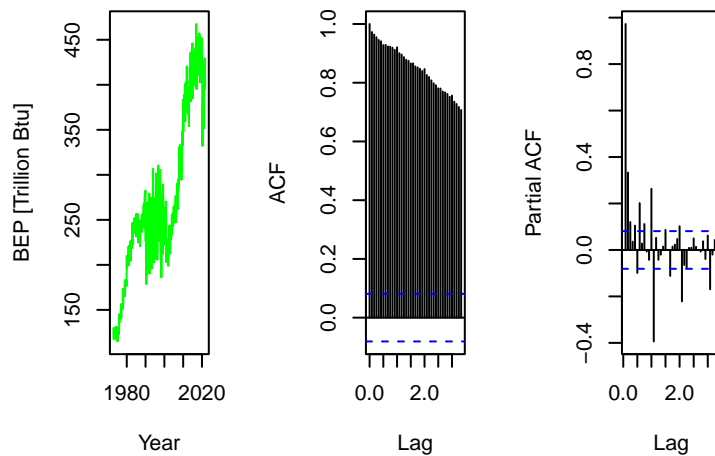
```
# Creating the time series object for each series
##Time series object of TBEP, TREP, HEPC
ts_us_TBEP <- ts(us_energy.df$TBEP, start = c(1973, 1), frequency= 12)
ts_us_TREP <- ts(us_energy.df$TREP, start = c(1973, 1), frequency= 12)
ts_us_HEPC <- ts(us_energy.df$HEPC, start = c(1973, 1), frequency= 12)
```

```

par(mfrow= c(1,3))
plot(ts_us_TBEP, type="l", col="green", xlab="Year", ylab="BEP [Trillion Btu]",
main="U.S.Total Biomass Energy Production (TBEP)")
acf(ts_us_TBEP, lag.max = 40, plot = TRUE)
pacf(ts_us_TBEP, lag.max = 40, plot = TRUE)

```

Biomass Energy Proc Series ts_us_TBE Series ts_us_TBE

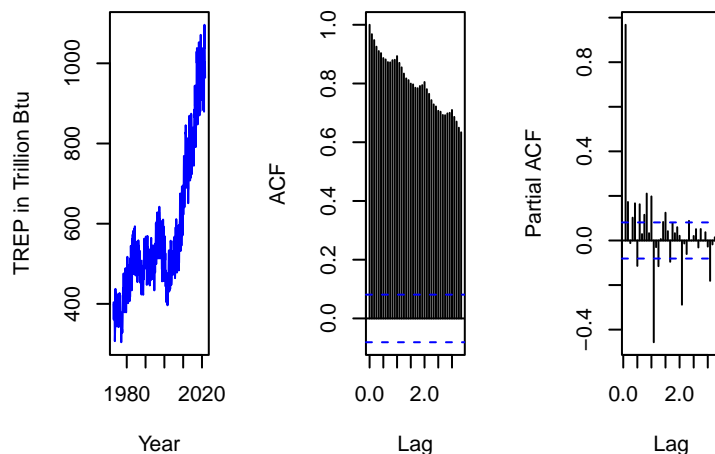


```

par(mfrow= c(1,3))
plot(ts_us_TREP, type="l", col="blue", xlab="Year", ylab="TREP in Trillion Btu", main="U.S. Total Renewable Energy Production (TREP)")
acf(ts_us_TREP, lag.max = 40, plot = TRUE)
pacf(ts_us_TREP, lag.max = 40, plot = TRUE)

```

Renewable Energy Proc Series ts_us_TRE Series ts_us_TRE

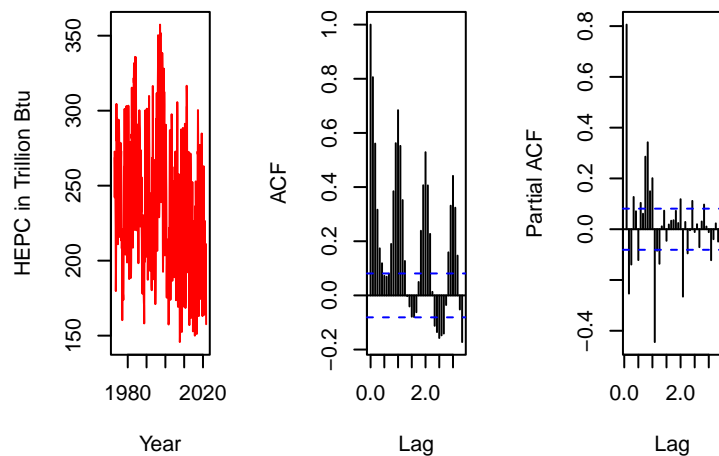


```

par(mfrow= c(1,3))
plot(ts_us_HEPC, type="l", col="red", xlab="Year", ylab="HEPC in Trillion Btu", main="U.S. Hydroelectric Production (HEPC)")
acf(ts_us_HEPC, lag.max = 40, plot = TRUE)
pacf(ts_us_HEPC, lag.max = 40, plot = TRUE)

```

Electric Power Consumption Series ts_us_HEP Series ts_us_HEP



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?

Yes, there is a trend in all the three variables (parameters). In general, the Total Biomass Energy Production (TBEP) and the Total Renewable Energy Production (TREP) show an increasing trend whereas the Hydroelectric Power Consumption show a decreasing trend over-time. Closely looking at the TBEP and TREP, the TBEP started showing a decrease at the final year of the observation period, however, the TREP shows a continued increase even at the end of the observation period.

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 Biomass Energy Production (TBEP) linear trend and regression
```

```
t <- c(1:585) # Creating vector 't'
```

```
linear_trend_TBEP = lm(ts_us_TBEP ~ t)
summary(linear_trend_TBEP)
```

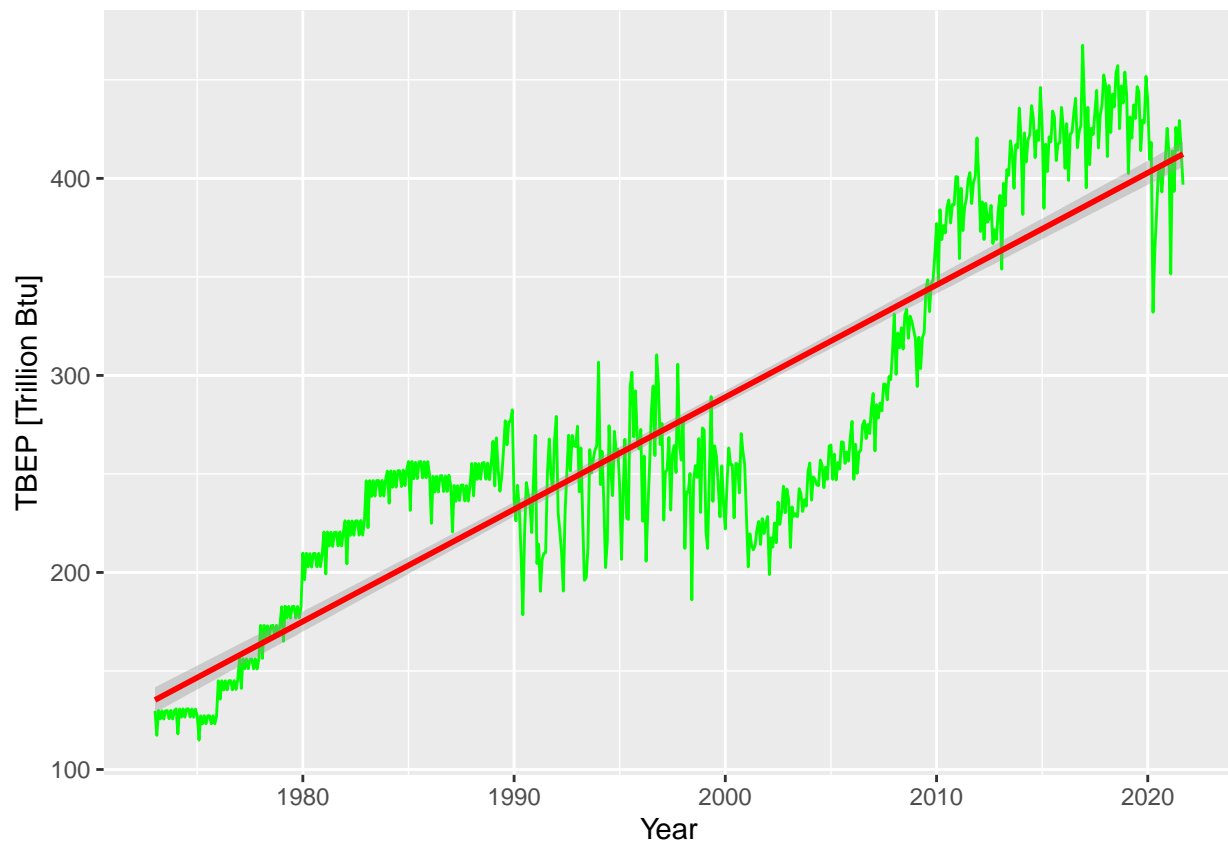
```
##
## Call:
## lm(formula = ts_us_TBEP ~ t)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -101.892  -24.306    4.932   33.103   82.292
##
```

```
## 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
```

```
beta0_TBEP <- as.numeric(linear_trend_TBEP$coefficients[1]) # Intercept
beta1_TBEP <- as.numeric(linear_trend_TBEP$coefficients[2]) # Slope

ggplot(us_energy.df, aes(x=my_date, y=us_energy.df[,2]))+
  geom_line(color="green")+
  ylab("TBEP [Trillion Btu]") +
  xlab("Year")+
  geom_smooth(color = "red", method = "lm")
```

```
## 'geom_smooth()' using formula 'y ~ x'
```



As time (t) increases by one unit, the TBEP increases by 0.474 times.

```
# Total Renewable Energy Production linear trend and regression
```

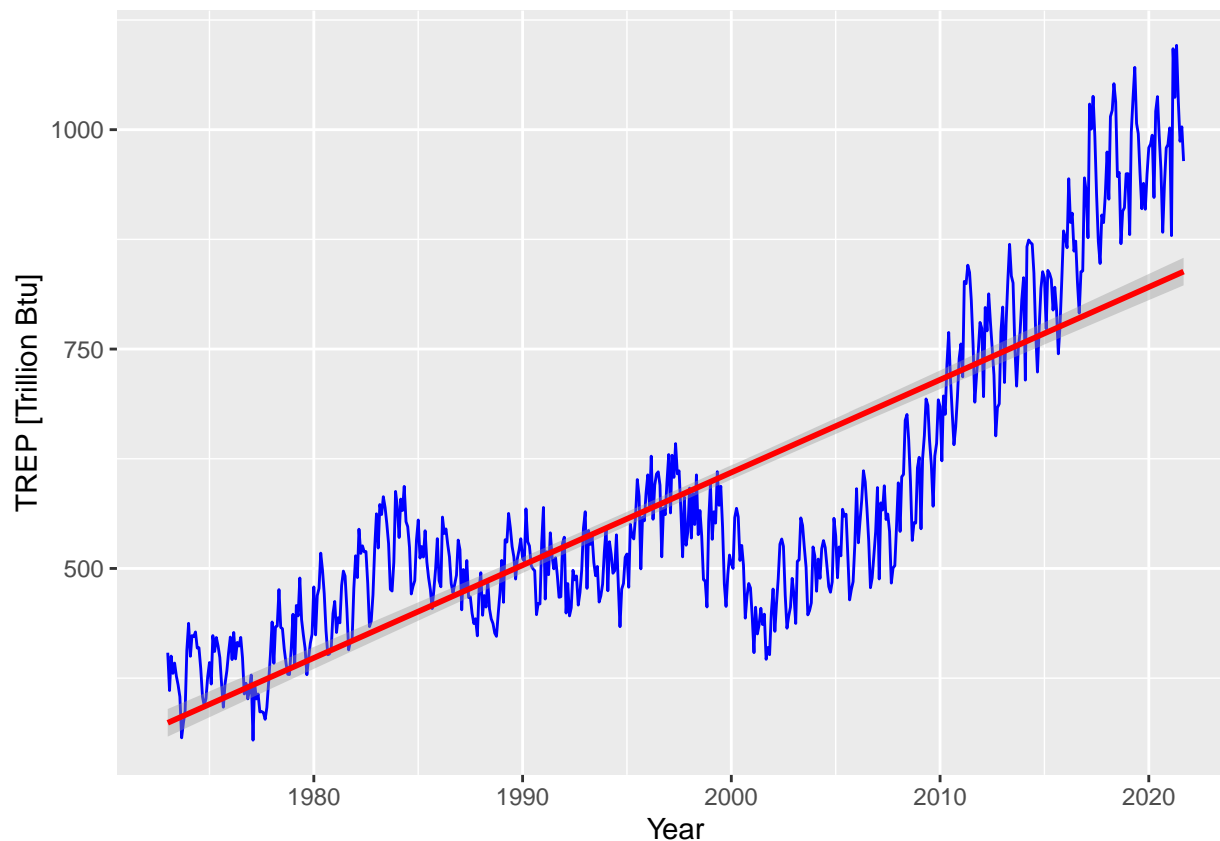
```
linear_trend_TREP = lm(ts_us_TREP ~ t)
summary(linear_trend_TREP)
```

```
##
## Call:
## lm(formula = ts_us_TREP ~ t)
##
## 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
```

```
beta0_TREP <- as.numeric(linear_trend_TREP$coefficients[1]) # Intercept
beta1_TREP <- as.numeric(linear_trend_TREP$coefficients[2]) # Slope
```

```
ggplot(us_energy.df, aes(x=my_date, y=us_energy.df[,3]))+
  geom_line(color="blue")+
  ylab("TREP [Trillion Btu]")+
  xlab("Year")+
  geom_smooth(color = "red", method = "lm")
```

```
## 'geom_smooth()' using formula 'y ~ x'
```

As time (t) increases by one unit, the TREP increases by 0.88 times.

```
#Hydroelectirc Power Consumption linear trend and regression
```

```
linear_trend_HEPC = lm(ts_us_HEPC ~ t)
summary(linear_trend_HEPC)
```

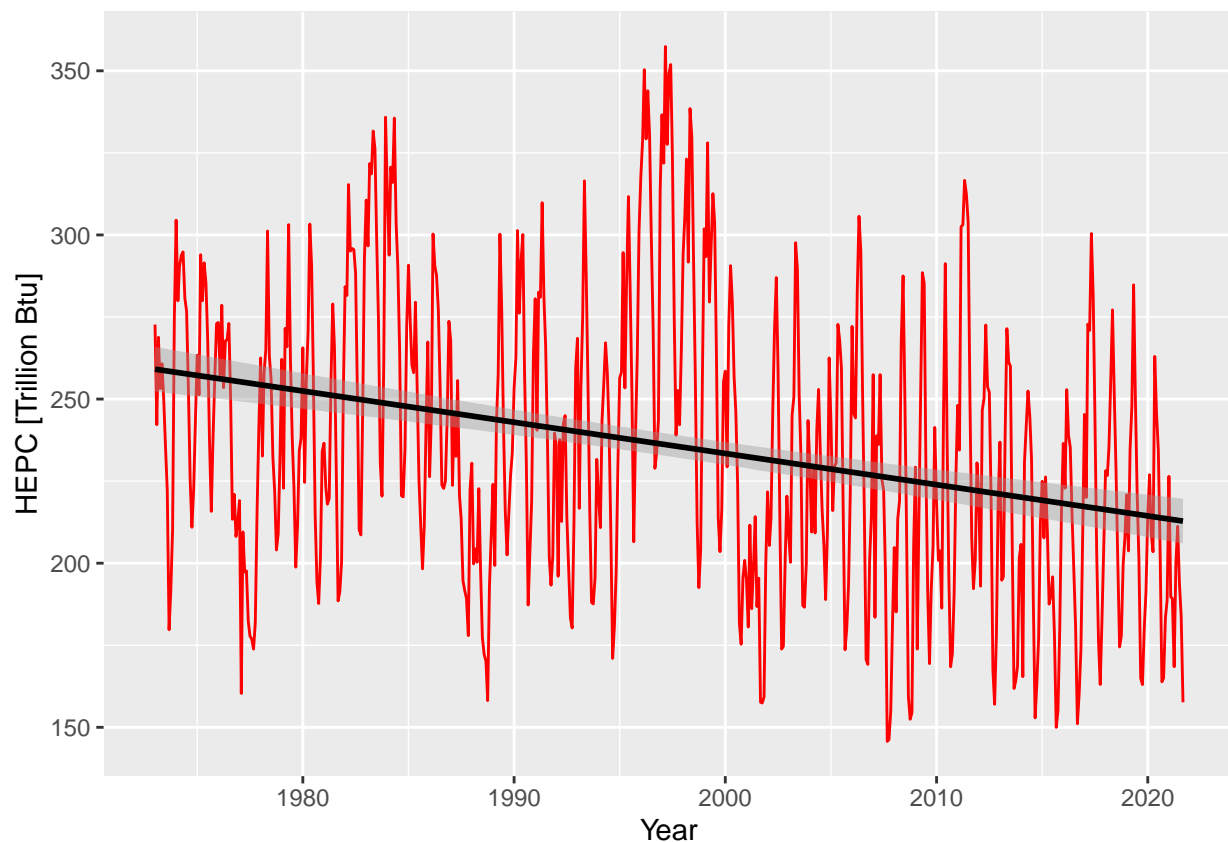
```
##
## Call:
## lm(formula = ts_us_HEPC ~ t)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -94.892 -31.300  -2.414   27.876  121.263
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  259.18303    3.47464   74.593  < 2e-16 ***
## t            -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
```

```
beta0_HEPC <- as.numeric(linear_trend_HEPC$coefficients[1]) # Intercept
beta1_HEPC <- as.numeric(linear_trend_HEPC$coefficients[2]) # Slope

ggplot(us_energy.df, aes(x=my_date, y=us_energy.df[,4]))+
  geom_line(color="red")+
  ylab("HEPC [Trillion Btu]") +
  xlab("Year")+
  geom_smooth(color = "black", method = "lm")
```

```
## 'geom_smooth()' using formula 'y ~ x'
```



As time (t) increases by one unit, the HEPC decreases by 0.079 times.

Q4

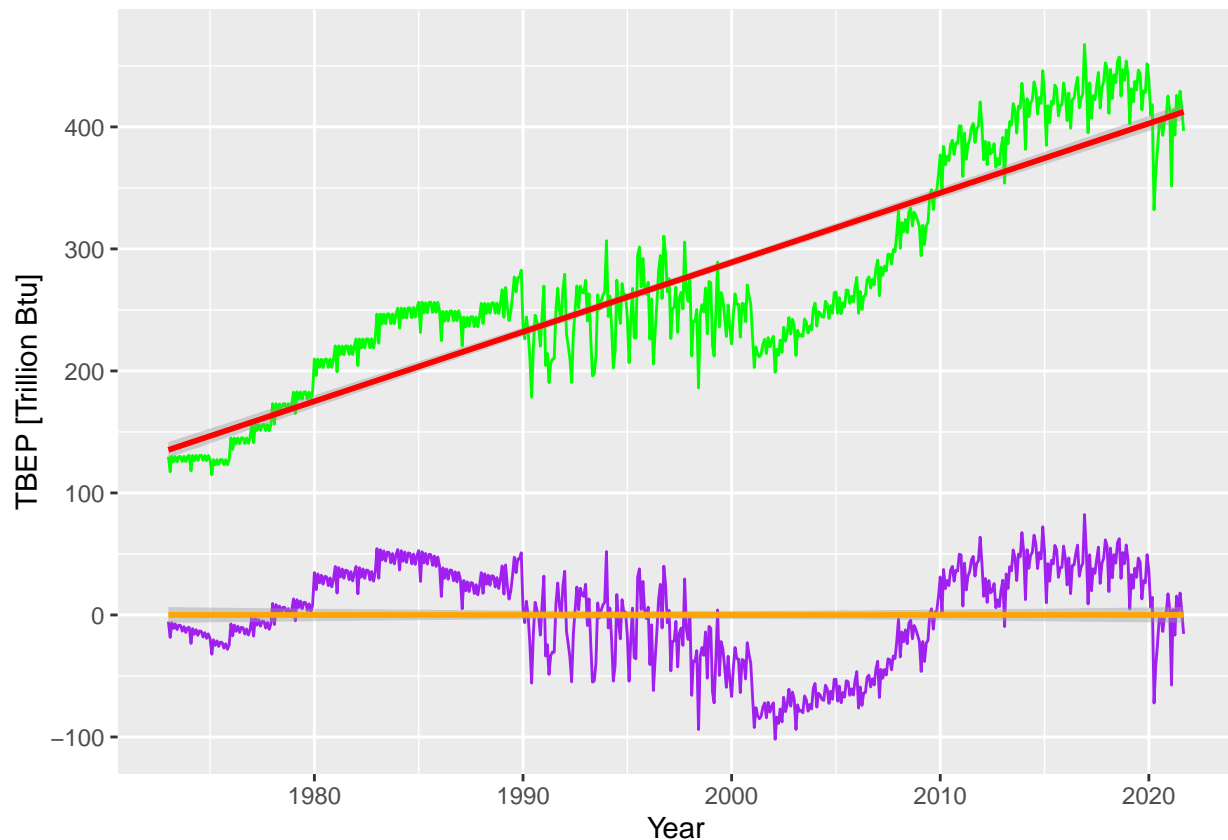
Use the regression coefficients from Q3 to detrend the series. Plot the detrended series and compare with the plots from Q1. What happened? Did anything change?

```
# Removing the trend from the TBEP series

detrend_ts_us_TBEP <- ts_us_TBEP - (beta0_TBEP + beta1_TBEP * t)
```

```
ggplot(us_energy.df, aes(x=my_date, y=us_energy.df[,2]))+
  geom_line(color="green")+
  ylab("TBEP [Trillion Btu]")+
  xlab("Year")+
  geom_smooth(color = "red", method = "lm")+
  geom_line(aes(y=detrend_ts_us_TBEP), color="purple")+
  geom_smooth(aes(y=detrend_ts_us_TBEP), color="orange", method="lm")
```

```
## 'geom_smooth()' using formula 'y ~ x'
## 'geom_smooth()' using formula 'y ~ x'
```



In the TBEP plot, the series in green is the original series with a red trend line that show a significantly increasing trend overtime. When we remove the trend from the original series, we get the new detrend series shown in purple with an orange trend line that has a zero slope. When the trend component removed from the series, the series seem to change to a kind of seasonal pattern.

```
# Removing the trend from the TREP series

detrend_ts_us_TREP <- ts_us_TREP - (beta0_TREP + beta1_TREP * t)

ggplot(us_energy.df, aes(x=my_date, y=us_energy.df[,3]))+
  geom_line(color="blue")
```

```

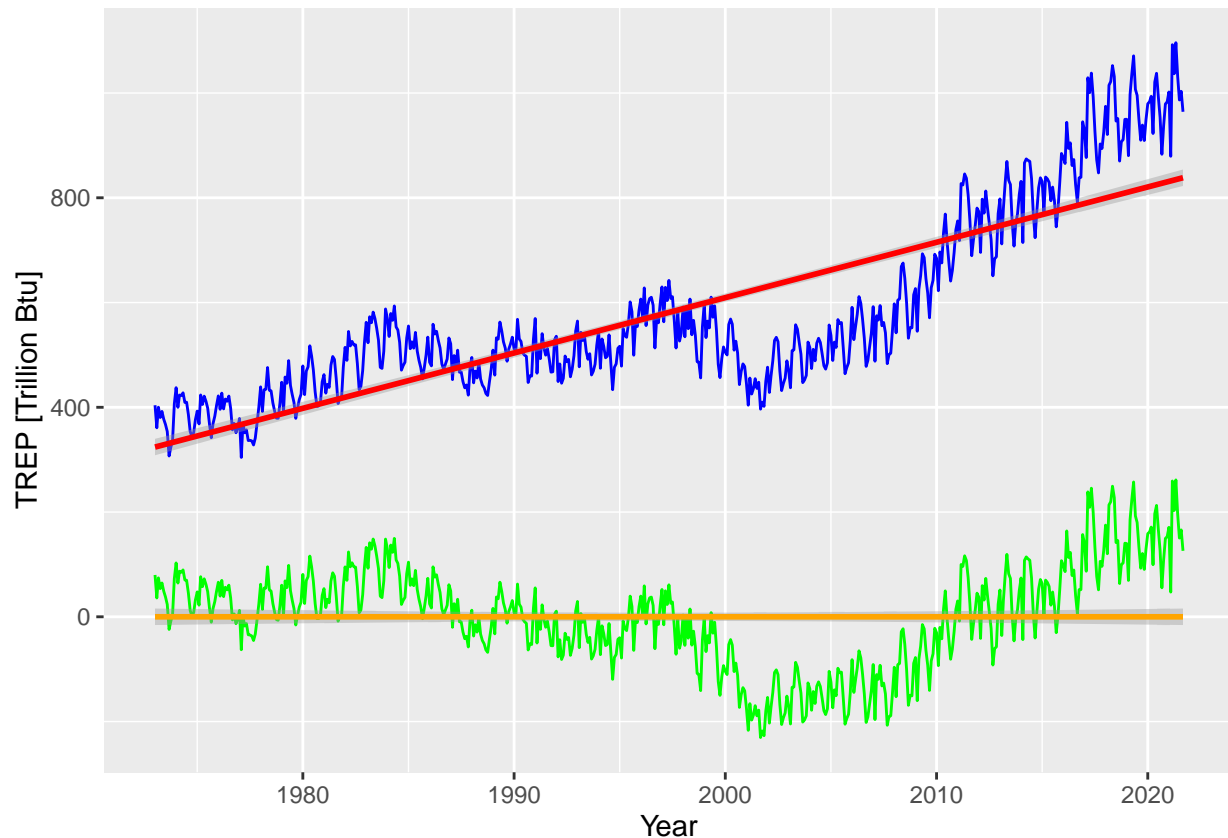
ylab("TREP [Trillion Btu]")+
xlab("Year")+
geom_smooth(color = "red", method = "lm")+
geom_line(aes(y=detrend_ts_us_TREP), color="green")+
geom_smooth(aes(y=detrend_ts_us_TREP), color="orange", method="lm")

```

```

## 'geom_smooth()' using formula 'y ~ x'
## 'geom_smooth()' using formula 'y ~ x'

```



In the TREP plot, the series in blue is the original series with a red trend line that show a significant increasing trend overtime. When we remove the trend from the original series, we get the new detrend series shown in green with an orange trend line that has a zero slope. When the trend component removed from the series, the series changed to a random pattern.

```

# Removing the trend from the HEPC series

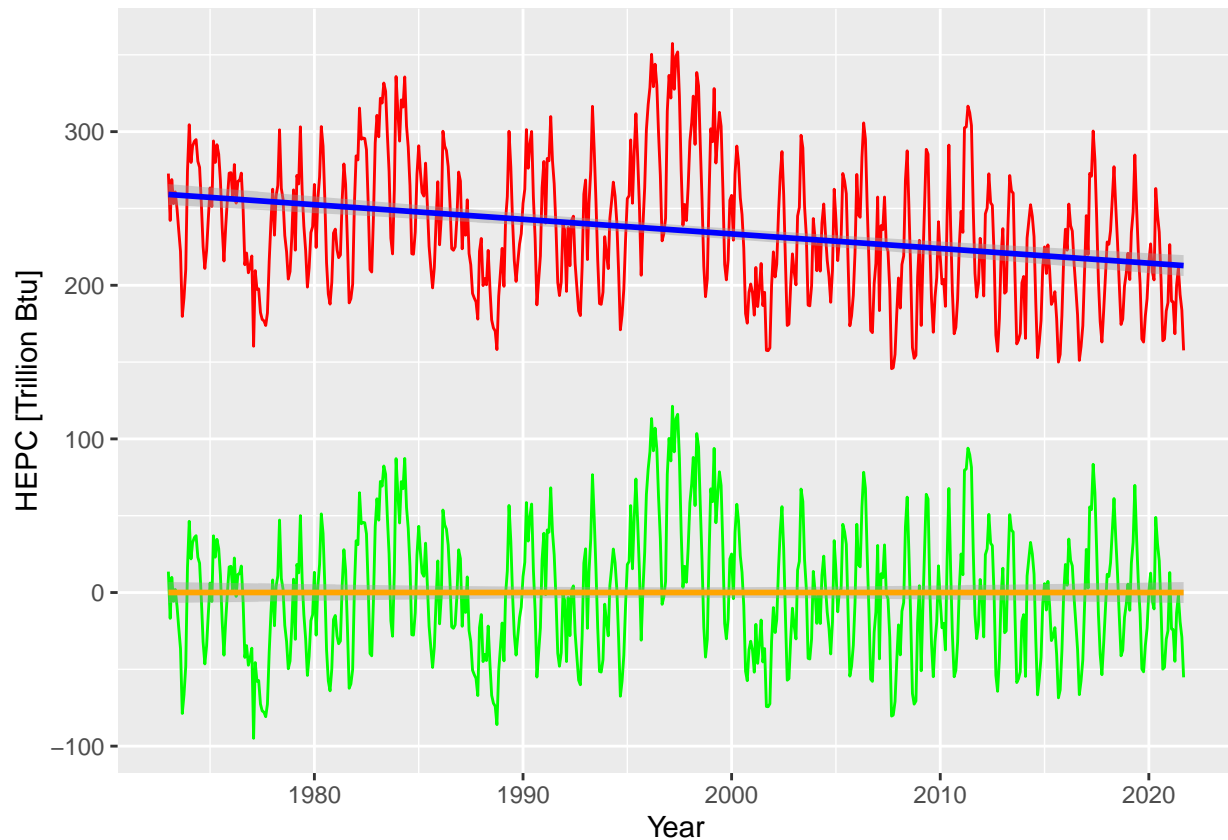
detrend_ts_us_HEPC <- ts_us_HEPC - (beta0_HEPC + beta1_HEPC * t)

ggplot(us_energy.df, aes(x=my_date, y=us_energy.df[,4]))+
  geom_line(color="red")+
  ylab("HEPC [Trillion Btu]")+
  xlab("Year")+
  geom_smooth(color = "blue", method = "lm")+

```

```
geom_line(aes(y=detrend_ts_us_HEPC), color="green")+
geom_smooth(aes(y=detrend_ts_us_HEPC), color="orange", method="lm")
```

```
## 'geom_smooth()' using formula 'y ~ x'
## 'geom_smooth()' using formula 'y ~ x'
```



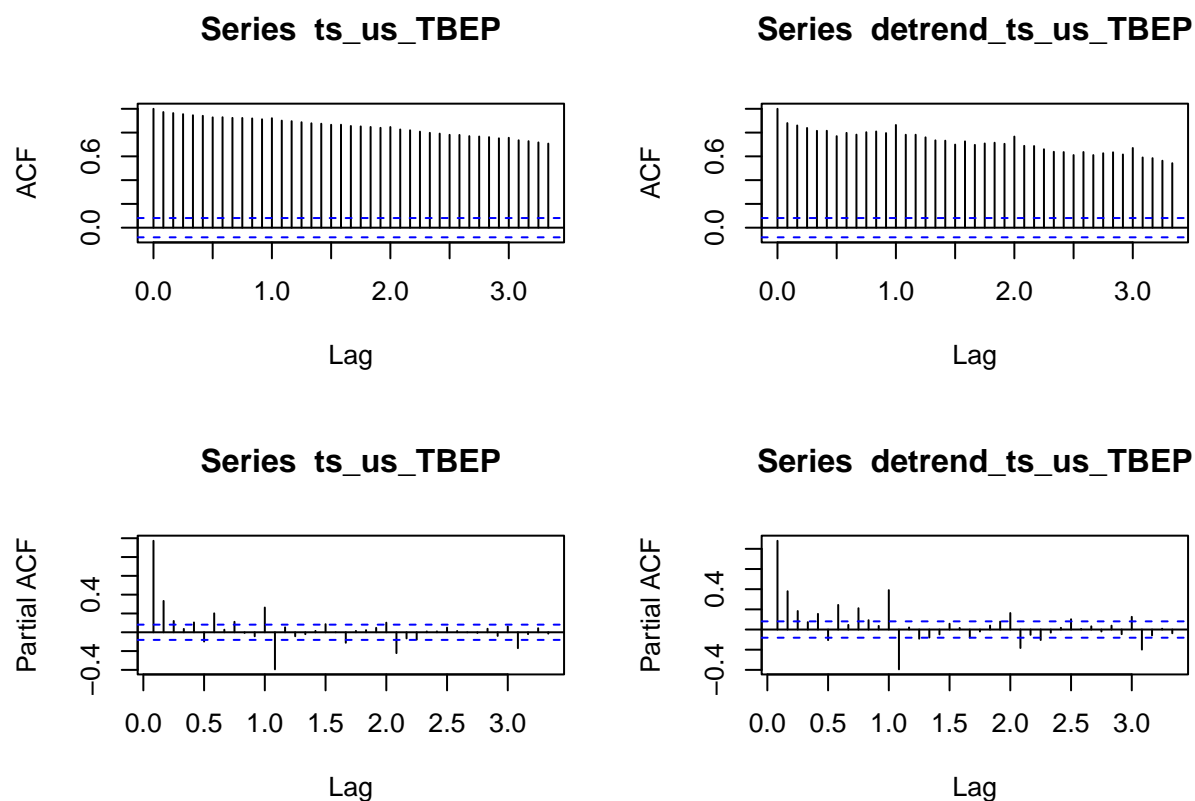
In the HEPC plot, the series in red is the original series with a blue trend line that show a decreasing trend overtime. When we remove the trend from the original series, we get the new detrend series shown in green with an orange trend line that has a zero slope. When the trend component removed from the series, the series changed to a kind of seasonal pattern.

Q5

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

```
# ACF and PACF of the detrended series for TBEP

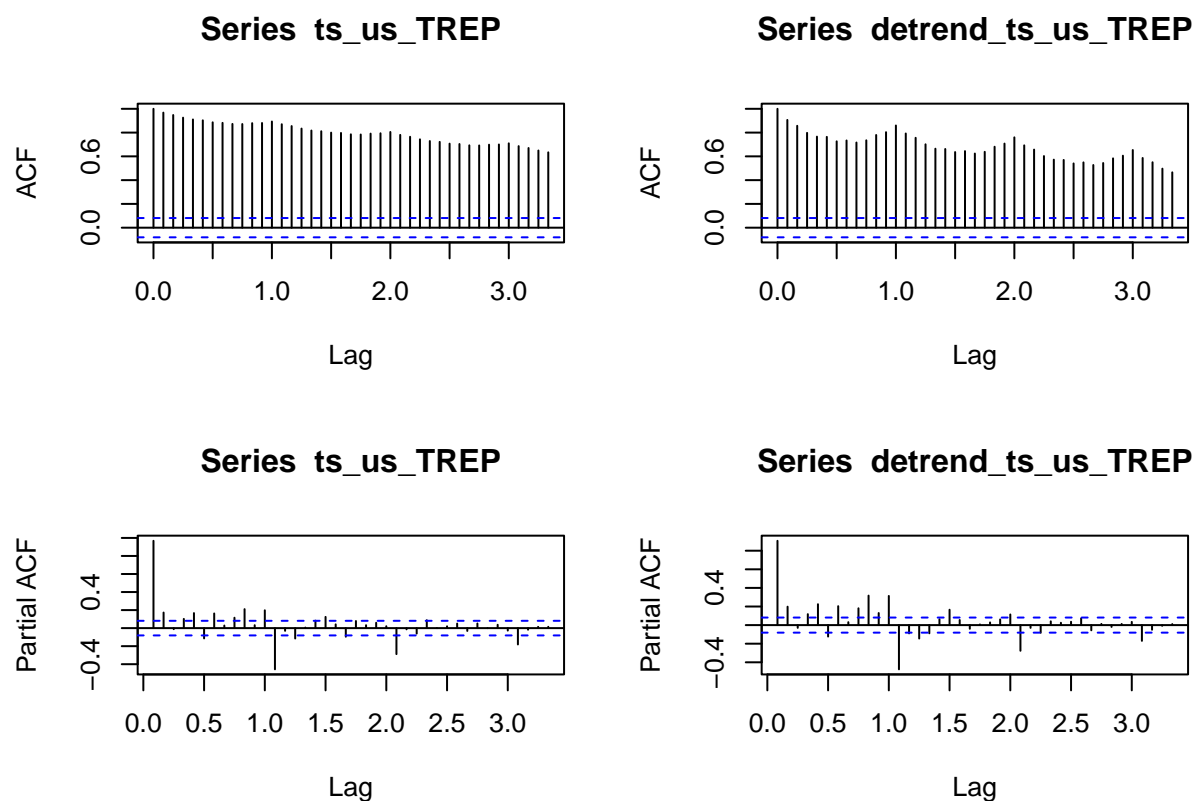
par(mfrow = c(2,2))
acf(ts_us_TBEP, lag.max = 40, plot = TRUE)
acf(detrend_ts_us_TBEP, lag.max = 40, plot=TRUE)
pacf(ts_us_TBEP, lag.max = 40, plot = TRUE)
pacf(detrend_ts_us_TBEP, lag.max = 40, plot = TRUE)
```



Yes, there is difference in the ACF and PACF plots of TBEP before and after the detrend. Before detrended, the ACF originally showed a slight decreasing trend overtime. When detrend, the ACF showed significant decreasing trend overtime. Similarly, in the case of PACF when detrended, there is a significant increasing trend in the correlation between the observations right after lag 1 to around lag 0.9 and at lag 2 which is higher than the original PACF.

```
# ACF and PACF of the detrended series for TREP

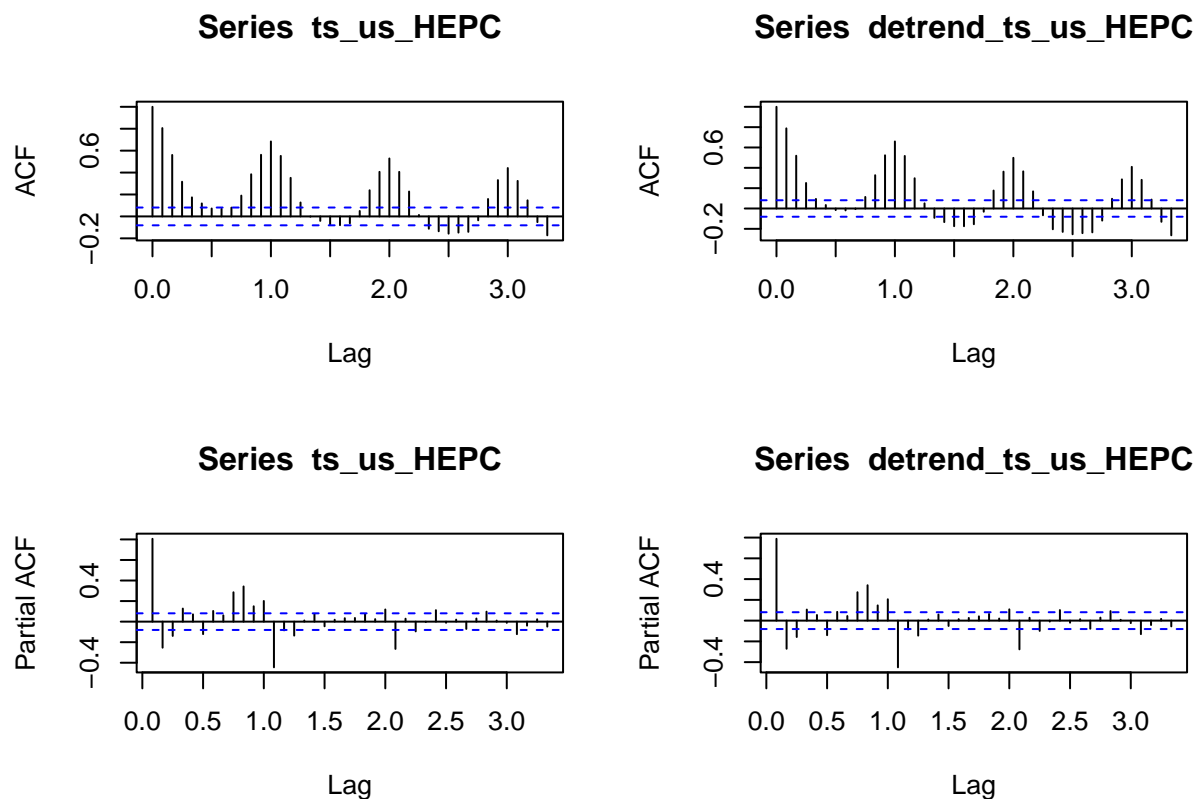
par(mfrow=c(2,2))
acf(ts_us_TREP, lag.max = 40, plot = TRUE)
acf(detrend_ts_us_TREP, lag.max = 40, plot=TRUE)
pacf(ts_us_TREP, lag.max = 40, plot = TRUE)
pacf(detrend_ts_us_TREP, lag.max = 40, plot = TRUE)
```



In the case of TREP, when detrended, the ACF plot show a seasonal decreasing pattern within a regular interval of about every lag. Similarly, in the PACF of the TREP, when detrended, the significance of the relationship among some observations around lags 0.9 and 1.5 get more pronounced than the trended series.

```
# ACF and PACF of the detrended series for HEPC

par(mfrow=c(2,2))
acf(ts_us_HEPC, lag.max = 40, plot = TRUE)
acf(detrend_ts_us_HEPC, lag.max = 40, plot=TRUE)
pacf(ts_us_HEPC, lag.max = 40, plot = TRUE)
pacf(detrend_ts_us_HEPC, lag.max = 40, plot = TRUE)
```



In the trended and detrended acf and pacf plots of the HEPC, there is only one difference that can be observed in the detrended ACF and that is an increase in the correlation of the observations overtime around lags 1.5 and 2.5 which was smaller in the original ACF.

Seasonal Component

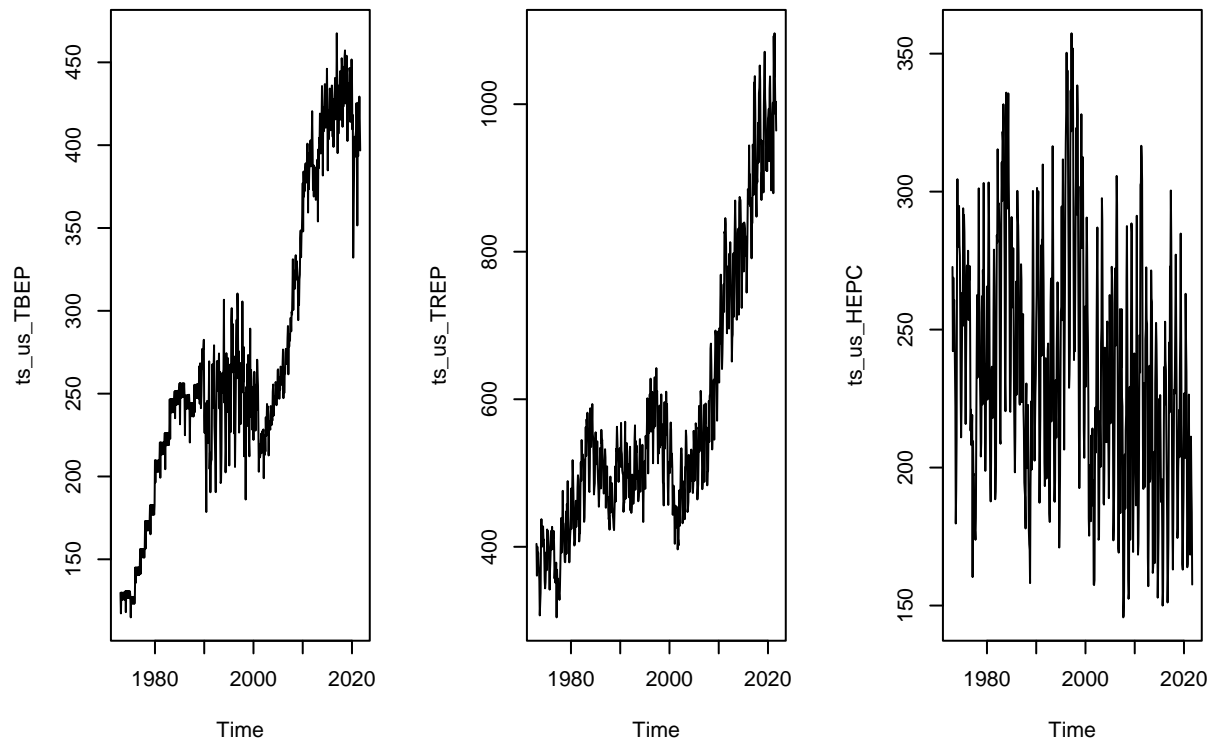
Set aside the detrended series and consider the original series again from Q1 to answer Q6 to Q8.

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.

```
##Time series object of TBEP, TREP, HEPC
ts_us_TBEP <- ts(us_energy.df$TBEP, start = c(1973, 1), frequency= 12)
ts_us_TREP <- ts(us_energy.df$TREP, start = c(1973, 1), frequency= 12)
ts_us_HEPC <- ts(us_energy.df$HEPC, start = c(1973, 1), frequency= 12)

par(mfrow=c(1,3))
plot(ts_us_TBEP)
plot(ts_us_TREP)
plot(ts_us_HEPC)
```

Looking at the time series plots the TBEP and HEPC series have a seasonal trend.

```
# TBEP
# Apply the seasonal means model to create the seasonal dummies

dummies_TBEP <- seasonaldummy(ts_us_TBEP)

# Then fit a linear model to the seasonal dummies

TBEP_seas_means_model=lm(us_energy.df$TBEP ~ dummies_TBEP)

summary(TBEP_seas_means_model)
```

```
##
## Call:
## lm(formula = us_energy.df$TBEP ~ dummies_TBEP)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -156.96  -51.40  -22.15   60.65  183.31
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    284.241    12.962   21.928  <2e-16 ***
```

```
## dummies_TBEPJan    -1.498    18.238  -0.082    0.9346
## dummies_TBEPFeb   -30.582    18.238  -1.677    0.0941 .
## dummies_TBEPMar    -8.873    18.238  -0.486    0.6268
## dummies_TBEPApr   -21.009    18.238  -1.152    0.2498
## dummies_TBEPMay   -14.065    18.238  -0.771    0.4409
## dummies_TBEPJun   -19.601    18.238  -1.075    0.2829
## dummies_TBEPJul    -3.499    18.238  -0.192    0.8479
## dummies_TBEPAug    -0.252    18.238  -0.014    0.9890
## dummies_TBEPSep   -12.518    18.238  -0.686    0.4928
## dummies_TBEPOct    -3.629    18.331  -0.198    0.8432
## dummies_TBEPNov    -9.592    18.331  -0.523    0.6010
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 89.81 on 573 degrees of freedom
## Multiple R-squared:  0.01056,    Adjusted R-squared:  -0.008439
## F-statistic: 0.5557 on 11 and 573 DF,  p-value: 0.8647
```

```
# Storing the regression coefficients
beta0_int_TBEP = TBEP_seas_means_model$coefficients[1]
beta1_coeff_TBEP = TBEP_seas_means_model$coefficients[2:12]
```

From the regression summary of the TBEP, the monthly TBEP showed a decrease throughout the year in a pattern of significant decrease and less significant decrease. For example, the decrease in the month of January is around 1.5 times followed by a significant decrease in the month of February by 30.58 times, and then a less significant decrease in March (8.87 times) followed by a significant decrease in April (by 14.07 times) and so on. Thus, in general the observation showed as time increases, the monthly TBEP decreases significantly and less significantly in an alternating manner.

```
# HEPC
# Apply the seasonal means model to create the seasonal dummies

dummies_HEPC <- seasonaldummy(ts_us_HEPC)

# Then fit a linear model to the seasonal dummies

HEPC_seas_means_model=lm(us_energy.df$HEPC ~ dummies_HEPC)

summary(HEPC_seas_means_model)
```

```
##
## Call:
## lm(formula = us_energy.df$HEPC ~ dummies_HEPC)
##
## 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 ***
```

```
## dummies_HEPCJan    13.558      6.883    1.970    0.04936 *
## dummies_HEPCFeb    -8.090      6.883   -1.175    0.24037
## dummies_HEPCMar    20.067      6.883    2.915    0.00369 **
## dummies_HEPCApr    16.619      6.883    2.414    0.01607 *
## dummies_HEPCMay    39.961      6.883    5.805    1.06e-08 ***
## dummies_HEPCJun    31.315      6.883    4.549    6.57e-06 ***
## dummies_HEPCJul     10.511      6.883    1.527    0.12732
## dummies_HEPCAug   -17.853      6.883   -2.594    0.00974 **
## dummies_HEPCSep   -49.852      6.883   -7.242    1.43e-12 ***
## dummies_HEPCOct   -48.086      6.919   -6.950    9.96e-12 ***
## dummies_HEPCNov   -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
```

```
# Storing the regression coefficients
beta0_int_HEPC = HEPC_seas_means_model$coefficients[1]
beta1_coeff_HEPC = HEPC_seas_means_model$coefficients[2:12]
```

From the regression summary, we can see that there is a decrease in HEPC during the months February (by 8.09), August (by 17.85), September (by 49.85), October (by 48.07) and November (by 32.19), with the highest decrease observed in September (by 49.85). In the remaining months, the series shows an increase in the HEPC by a range of 10.51 - 39.96, except the month of February which show a decrease in the HEPC (by 8.090). The highest increase is in the month of May (by 39.96) and the lowest increase is in the month of July (by 10.51). These in general created a pattern of HEPC increase in some months (from March to July) and then decrease in consumption in the following months (from August to November).The exceptional observation (the decrease in the month of February) requires additional information to explain.

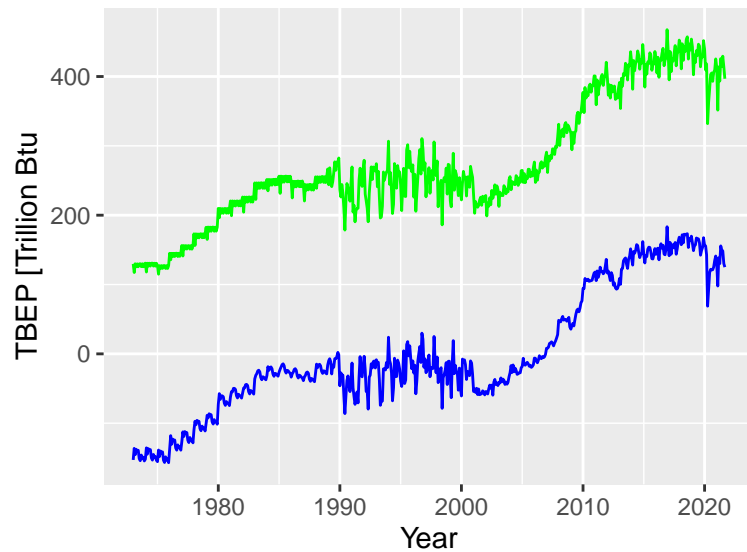
Q7

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?

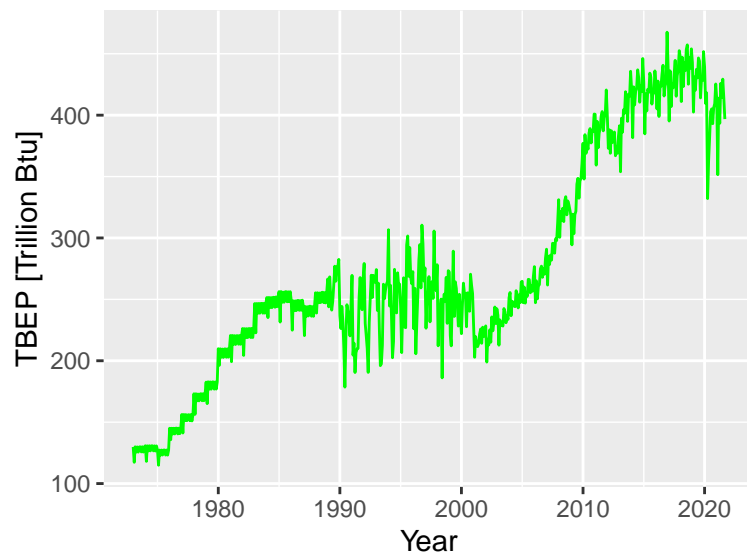
```
# For TBEP
# Compute the seasonal component
ts_us_TBEP_seas_comp=array(0,585)
for(i in 1:585)
{
  ts_us_TBEP_seas_comp[i]=(beta0_int_TBEP+beta1_coeff_TBEP%*%dummies_HEPC[i,])
}
# Removing the seasonal component
deseason_ts_us_TBEP <- us_energy.df$TBEP - ts_us_TBEP_seas_comp

par(mfrow=c(1,2))
ggplot(us_energy.df, aes(x=my_date, y=TBEP)) +
  geom_line(color="green") +
  ylab("TBEP [Trillion Btu]") +
```

```
xlab("Year") +  
geom_line(aes(y=deseason_ts_us_TBEP), col="blue")
```



```
ggplot(us_energy.df, aes(x=my_date, y=us_energy.df[,2]))+  
  geom_line(color="green")+  
  ylab("TBEP [Trillion Btu]")+  
  xlab("Year")
```



When compared with the original series, there is a small change in the pattern. When deseasoned, the series loses its seasonal pattern and demonstrates a random movement overtime.

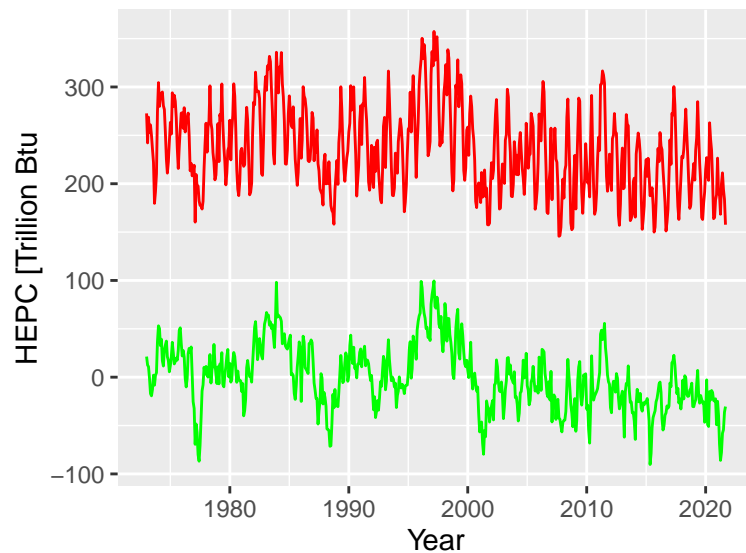
```
# For HEPC  
# Compute the seasonal component
```

```

ts_us_HEPC_seas_comp=array(0,585)
for(i in 1:585)
{
  ts_us_HEPC_seas_comp[i]=(beta0_int_HEPC+beta1_coeff_HEPC%%dummies_HEPC[i,])
}
# Removing the seasonal component
deseason_ts_us_HEPC <- us_energy.df$HEPC - ts_us_HEPC_seas_comp

par(mfrow=c(1,2))
ggplot(us_energy.df, aes(x=my_date, y=HEPC)) +
  geom_line(color="red") +
  ylab("HEPC [Trillion Btu]") +
  xlab("Year") +
  geom_line(aes(y=deseason_ts_us_HEPC), col="green")

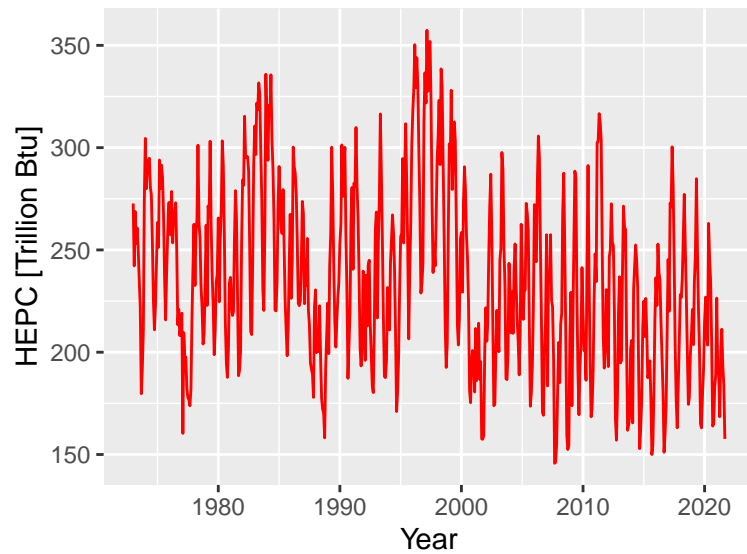
```



```

ggplot(us_energy.df, aes(x=my_date, y=us_energy.df[,4]))+
  geom_line(color="red")+
  ylab("HEPC [Trillion Btu]") +
  xlab("Year")

```



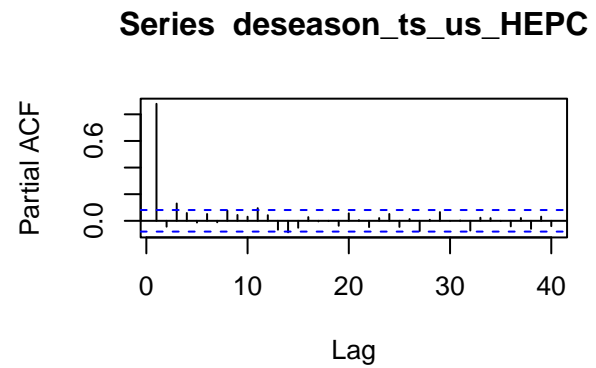
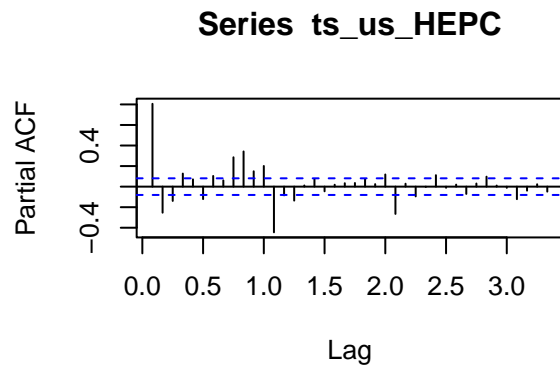
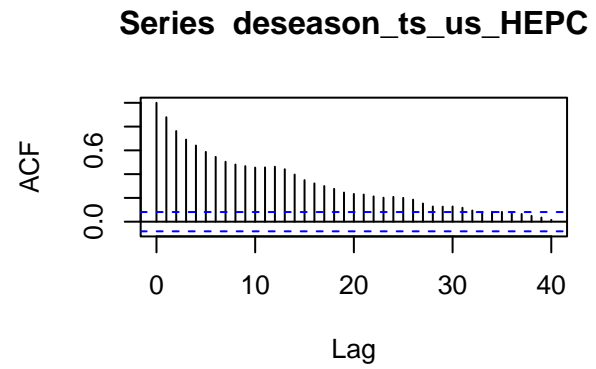
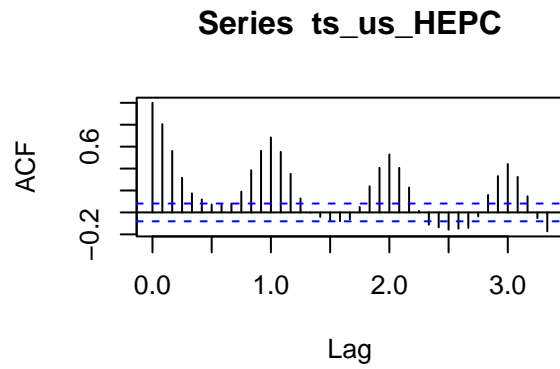
When compared with the original series, yes, there is a change in the pattern. When deseasoned, the series loses its seasonal pattern and demonstrates a random movement overtime.

Q8

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

```
# Comparing the acf and pacf plots of seasonal and deseason series of HEPC

par(mfrow=c(2,2))
acf(ts_us_HEPC, lag.max = 40, plot = TRUE)
acf(deseason_ts_us_HEPC, lag.max = 40)
pacf(ts_us_HEPC, lag.max = 40, plot = TRUE)
pacf(deseason_ts_us_HEPC, lag.max = 40)
```



When comparing the ACF plots of the HEPC, there is a completely different pattern change from seasonal increase-decrease (in the original ACF) to a continuous gradual significant decrease as the lags increase reaching to insignificant correlation among the observations at lag 40. In case of the PACF, when the seasonal component removed, the significant correlation of observations after the first lag disappear along with it leaving the significant at lag one unchanged/ the same.