ENV 790.30 - Time Series Analysis for Energy Data | Spring 2025

Assignment 3 - Due date 02/04/25

Jessalyn Chuang

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_A03_Sp25.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 2024 **Monthly** Energy Review. Once again you will work only with the following columns: Total Renewable Energy Production and Hydroelectric Power Consumption. Create a data frame structure with these two 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)
library(ggplot2)
library(forecast)
library(Kendall)
library(tseries)
library(readxl)
library(cowplot)
```

##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)

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

```
## New names:
## * `` -> `...1`
## * `` -> `...2`
## * `` -> `...3`
## * `` -> `...4`
## * `` -> `...5`
## * `` -> `...6`
## * `` -> `...7`
## * `` -> `...8`
## * `` -> `...9`
## * `` -> `...10`
## * `` -> `...11`
## * `` -> `...12`
## * `` -> `...13`
## * `` -> `...14`
```

```
#Extract column names from from row 11
read_col_names <- read_excel(path="./Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_Sou
```

```
## New names:
## * `` -> `...1`
## * `` -> `...2`
## * `` -> `...3`
## * `` -> `...4`
## * `` -> `...5`
## * `` -> `...6`
## * `` -> `...7`
## * `` -> `...8`
## * `` -> `...9`
## * `` -> `...10`
## * `` -> `...11`
## * `` -> `...12`
## * `` -> `...13`
## * `` -> `...14`
```

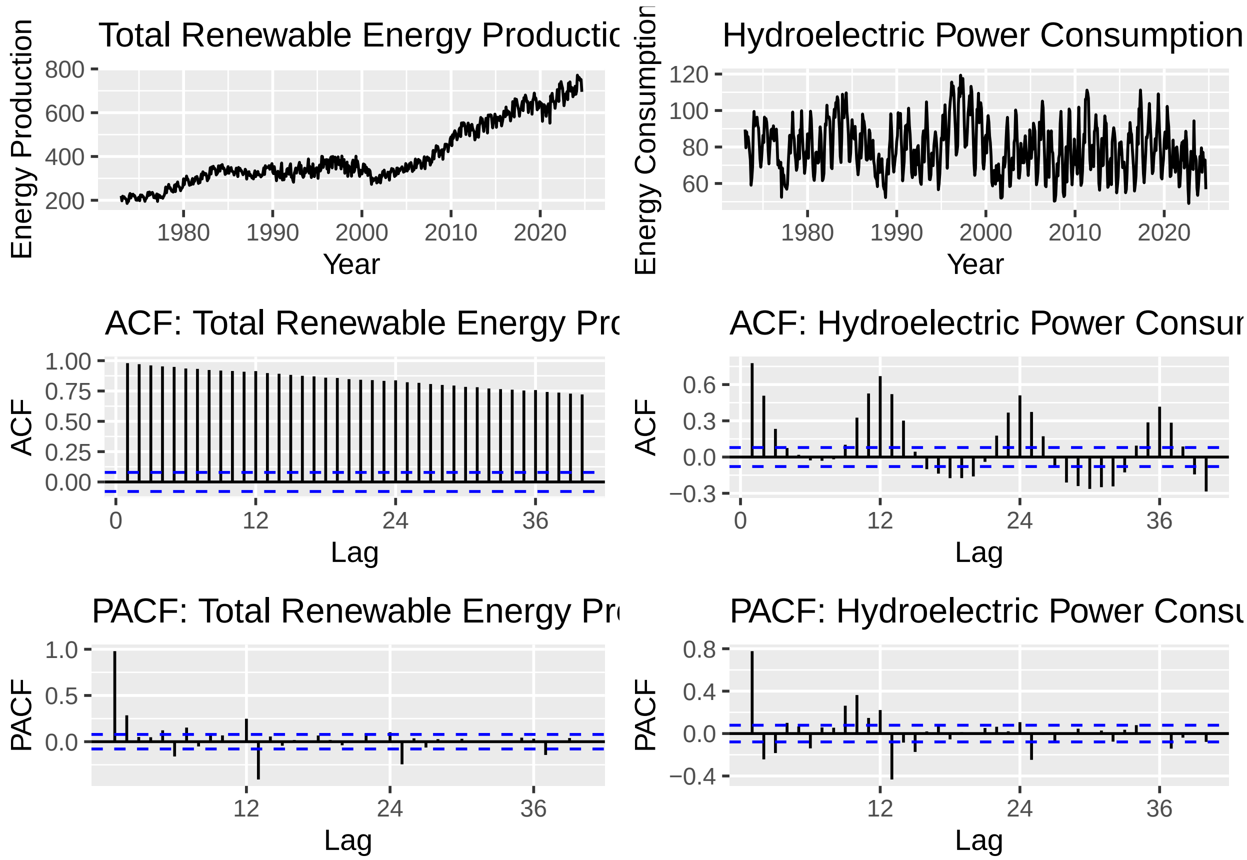
```
#Assign the column names to the data set
colnames(energy_data) <- read_col_names

#Visualize the first rows of the data set
head(energy_data)
```

```
## # A tibble: 6 x 14
##   Month                `Wood Energy Production` `Biofuels Production`
##   <dttm>                                  <dbl> <chr>
```

```
## 1 1973-01-01 00:00:00                    130. Not Available
## 2 1973-02-01 00:00:00                    117. Not Available
## 3 1973-03-01 00:00:00                    130. Not Available
## 4 1973-04-01 00:00:00                    125. Not Available
## 5 1973-05-01 00:00:00                    130. Not Available
## 6 1973-06-01 00:00:00                    125. Not Available
## # i 11 more variables: `Total Biomass Energy Production` <dbl>,
## #   `Total Renewable Energy Production` <dbl>,
## #   `Hydroelectric Power Consumption` <dbl>,
## #   `Geothermal Energy Consumption` <dbl>, `Solar Energy Consumption` <chr>,
## #   `Wind Energy Consumption` <chr>, `Wood Energy Consumption` <dbl>,
## #   `Waste Energy Consumption` <dbl>, `Biofuels Consumption` <chr>,
## #   `Total Biomass Energy Consumption` <dbl>, ...
```

```r
#pulling only Renewable Energy Production and Hydroelectric Consumption
energy_subset <- energy_data[, 5:6]

#turning into a time series
energy_subset_ts <- ts(energy_subset[,1:2],start=c(1973,1),frequency=12)
```

```r
#time series
ts_renewable <- autoplot(energy_subset_ts[,1], main = colnames(energy_subset_ts)[1],
                         xlab = "Year",ylab = "Energy Production")
ts_hydro <- autoplot(energy_subset_ts[,2], main = colnames(energy_subset_ts)[2],
                     xlab = "Year", ylab = "Energy Consumption")

# ACF and PACF using ggAcf() and ggPacf()
acf_renewable <- ggAcf(energy_subset_ts[,1], lag.max = 40) +
  ggtitle(paste("ACF:", colnames(energy_subset_ts)[1]))
acf_hydro <- ggAcf(energy_subset_ts[,2], lag.max = 40) +
  ggtitle(paste("ACF:", colnames(energy_subset_ts)[2]))

pacf_renewable <- ggPacf(energy_subset_ts[,1], lag.max = 40) +
  ggtitle(paste("PACF:", colnames(energy_subset_ts)[1]))
pacf_hydro <- ggPacf(energy_subset_ts[,2], lag.max = 40) +
  ggtitle(paste("PACF:", colnames(energy_subset_ts)[2]))

# Using plot_grid to arrange plots (all are now ggplot objects)
plot_grid(ts_renewable, ts_hydro, acf_renewable, acf_hydro, pacf_renewable,
          pacf_hydro, ncol = 2)
```



Q2

From the plot in Q1, do the series Total Renewable Energy Production and Hydroelectric Power Consumption appear to have a trend? If yes, what kind of trend?

Total Renewable Energy Production seems to have a clear long-term upward trend. This time series plot shows a steady increase over time, suggesting that renewable energy production has been growing consistently. The ACF plot confirms this trend since the auto-correlations remain strong for many lags, which is characteristic of a non-stationary series with a persistent upward movement. The PACF plot also indicates that past values strongly influence future values.

Hydroelectric power consumption does not have a clear long-term upward or downward trend. It instead fluctuates around a relatively stable mean, suggesting stationarity in the mean but with seasonality present. This is shown in the time series plot, which shows repetition in peaks and troughs over time. The ACF plot supports this since it shows significant autocorrelations at particular lags, indicating a recurring seasonal pattern. The PACF further supports this by showing notable spikes, suggesting taht hydroelectric consumption depends on past seasonal patterns.

Q3

Use the *lm()* function to fit a linear trend to the two 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.

```r
nobs <- nrow(energy_subset)
col_renewable <- 1
col_hydro <- 2
#Create vector t
t <- c(1:nobs)

#Fit a linear trend to Renewable Energy Production
renewable_linear_model <- lm(energy_subset[[col_renewable]] ~ t)
summary(renewable_linear_model)
```

```
##
## Call:
## lm(formula = energy_subset[[col_renewable]] ~ t)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -151.11  -37.84   13.53   41.76  149.42
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 176.87293    4.96189   35.65   <2e-16 ***
## t             0.72393    0.01382   52.37   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 61.75 on 619 degrees of freedom
## Multiple R-squared:  0.8159, Adjusted R-squared:  0.8156
## F-statistic:  2743 on 1 and 619 DF,  p-value: < 2.2e-16
```

```r
renewable_beta0 <- as.numeric(renewable_linear_model$coefficients[1])
renewable_beta1 <- as.numeric(renewable_linear_model$coefficients[2])
```

For Renewable Energy Production, the intercept of 176.87 is the estimated renewable energy production at the start of the dataset when $t = 0$. The slope being 0.7239 indicates that renewable energy is increasing by this many units per time step. Both the intercept and time coefficient are highly significant, meaning that the null hypothesis that there is no linear relationship between time and this seris can be rejected. There is indeed a strong and significant upward trend in renewable energy production over time.

```r
#Fit a linear trend to Hydroelectric Power Consumption
hydro_linear_model <- lm(energy_subset[[col_hydro]] ~ t)
summary(hydro_linear_model)
```

```
##
## Call:
## lm(formula = energy_subset[[col_hydro]] ~ t)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -29.995 -10.422  -0.720   9.161  39.624
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) 82.96766    1.12339  73.855  < 2e-16 ***
## t           -0.01098    0.00313  -3.508 0.000485 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.98 on 619 degrees of freedom
## Multiple R-squared:  0.01949,    Adjusted R-squared:  0.01791
## F-statistic:  12.3 on 1 and 619 DF,  p-value: 0.0004848
```

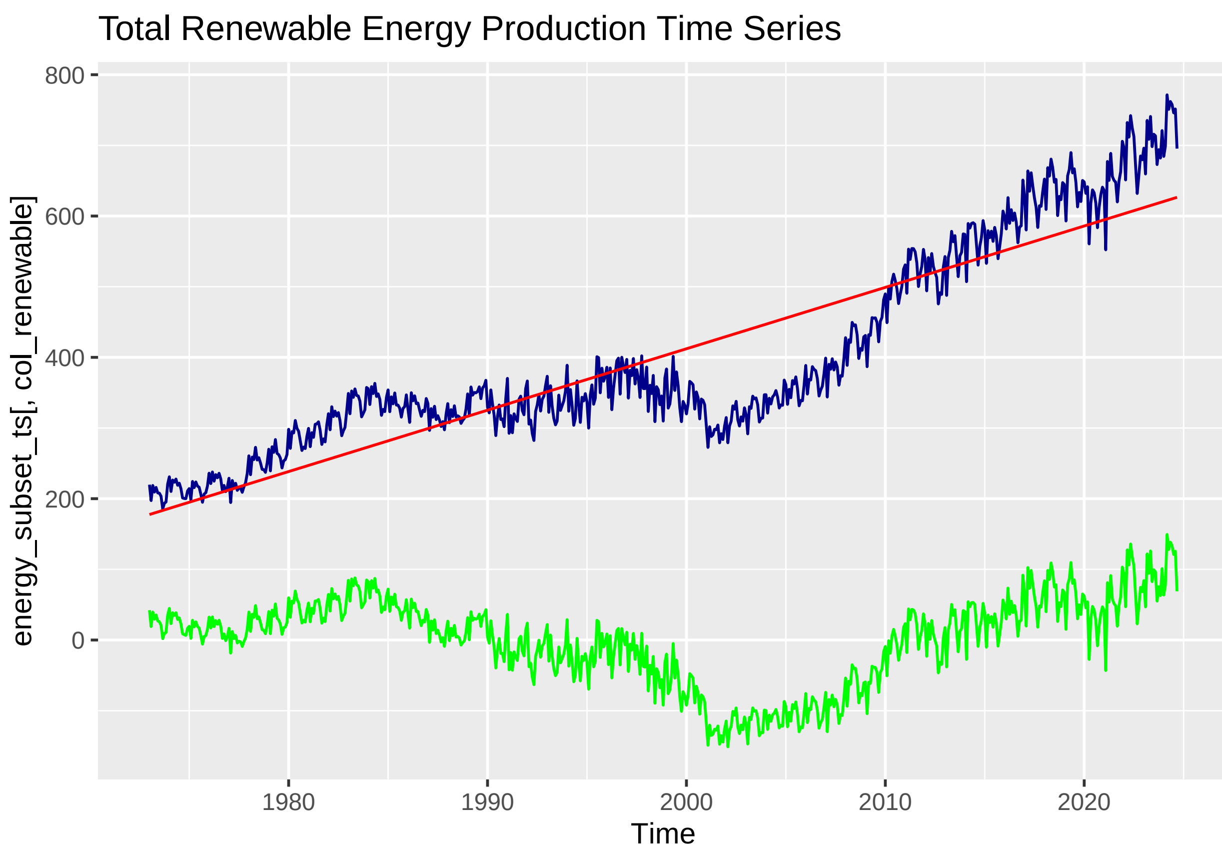
```
hydro_beta0 <- as.numeric(hydro_linear_model$coefficients[1])
hydro_beta1 <- as.numeric(hydro_linear_model$coefficients[2])
```

The intercept of 82.97 represents the estimated hydroelectric power consumption at the start of the dataset. The slop of -0.01098 being so small suggests that there is just a very small decrease over time. However, since the intercept and slope coefficients have p-values that are much smaller than 0.05, we can still reject the null hypothesis and conclude that time does have a statistically significant effect on hydroelectric power consumption. The low R-squared value (0.01949) suggests that while there is a trend, time alone does not explain much of the variability in hydro consumption.

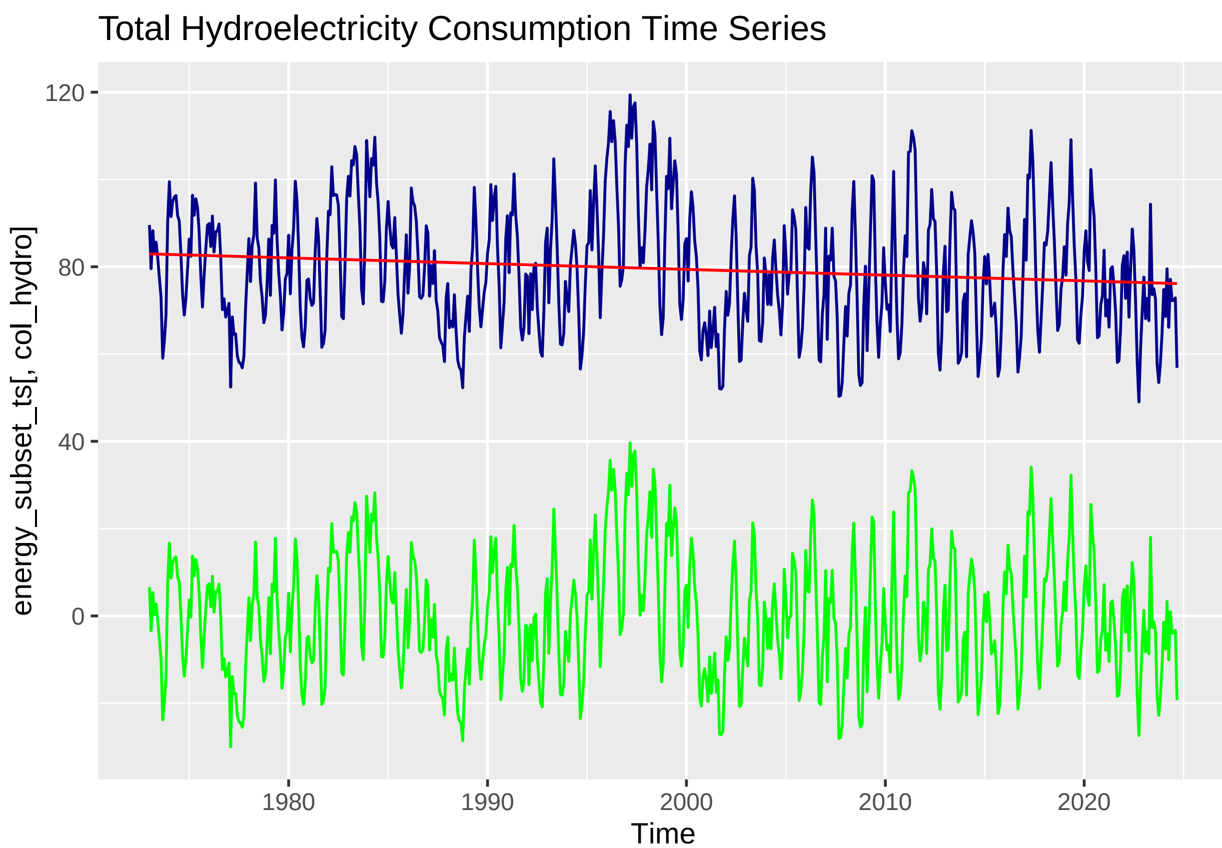
Q4

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?

```
renewable_linear_trend <- renewable_beta0 + renewable_beta1 * t
ts_renewable_linear <- ts(renewable_linear_trend,star=c(1973,1),frequency=12)

detrend_renewable <- energy_subset[,col_renewable] - renewable_linear_trend
ts_renewable_detrend <- ts(detrend_renewable, start = c(1973,1),frequency = 12)

#Plot
autoplot(energy_subset_ts[,col_renewable],color="darkblue")+
  autolayer(ts_renewable_detrend,series="Detrended",color="green")+
  autolayer(ts_renewable_linear,series="Linear Component",color="red")+
  labs(title = "Total Renewable Energy Production Time Series")
```



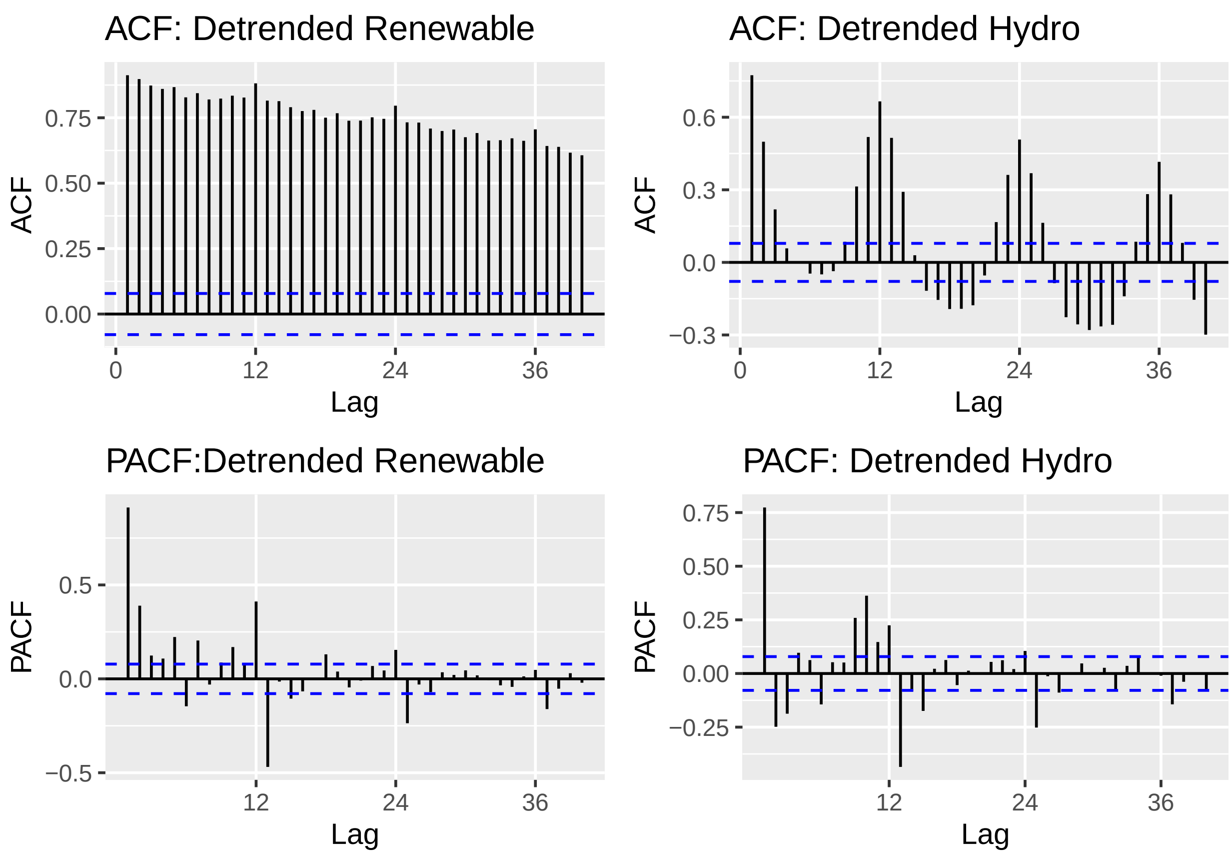

```
hydro_linear_trend <- hydro_beta0 + hydro_beta1 * t
ts_hydro_linear <- ts(hydro_linear_trend,start=c(1973,1),frequency=12)

detrend_hydro <- energy_subset[,col_hydro] - hydro_linear_trend
ts_hydro_detrend <- ts(detrend_hydro, start = c(1973,1),frequency = 12)

#Plot
autoplot(energy_subset_ts[,col_hydro],color="darkblue")+
  autolayer(ts_hydro_detrend,series="Detrended",color="green")+
  autolayer(ts_hydro_linear,series="Linear Component",color="red")+
  labs(title = "Total Hydroelectricity Consumption Time Series")
```



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, but not mandatory. Did the plots change? How?

```r
# ACF and PACF using ggAcf() and ggPacf()
acf_renewable_detrend <- ggAcf(ts_renewable_detrend, lag.max = 40) +
  ggtitle(paste("ACF: Detrended Renewable"))
acf_hydro_detrend <- ggAcf(ts_hydro_detrend, lag.max = 40) +
  ggtitle(paste("ACF: Detrended Hydro"))

pacf_renewable_detrend <- ggPacf(ts_renewable_detrend, lag.max = 40) +
  ggtitle(paste("PACF:Detrended Renewable"))
pacf_hydro_detrend <- ggPacf(ts_hydro_detrend, lag.max = 40) +
  ggtitle(paste("PACF: Detrended Hydro"))

# Using plot_grid to arrange plots (all are now ggplot objects)
plot_grid(acf_renewable_detrend, acf_hydro_detrend, pacf_renewable_detrend,
          pacf_hydro_detrend, ncol = 2)
```



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.

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_renewable <- seasonaldummy(ts_renewable_detrend)

seas_means_model_renewable <- lm(ts_renewable_detrend ~ dummies_renewable)
summary(seas_means_model_renewable)
```

```
##
## Call:
## lm(formula = ts_renewable_detrend ~ dummies_renewable)
```

```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -149.18  -38.16   14.42   41.50  134.67
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)             7.858      8.504   0.924  0.35584
## dummies_renewableJan    5.592     11.968   0.467  0.64048
## dummies_renewableFeb  -31.452     11.968  -2.628  0.00881 **
## dummies_renewableMar    6.892     11.968   0.576  0.56491
## dummies_renewableApr   -6.449     11.968  -0.539  0.59023
## dummies_renewableMay    7.923     11.968   0.662  0.50822
## dummies_renewableJun   -3.394     11.968  -0.284  0.77682
## dummies_renewableJul    2.126     11.968   0.178  0.85906
## dummies_renewableAug   -5.878     11.968  -0.491  0.62351
## dummies_renewableSep  -31.209     11.968  -2.608  0.00934 **
## dummies_renewableOct  -18.757     12.026  -1.560  0.11937
## dummies_renewableNov  -19.982     12.026  -1.661  0.09713 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 60.73 on 609 degrees of freedom
## Multiple R-squared:  0.04839,    Adjusted R-squared:  0.0312
## F-statistic: 2.815 on 11 and 609 DF,  p-value: 0.001358
```

```
beta_intercept_renewable <- seas_means_model_renewable$coefficients[1]
beta_coeff_renewable <- seas_means_model_renewable$coefficients[2:12]
```

```
dummies_hydro <- seasonaldummy(ts_hydro_detrend)

seas_means_model_hydro <- lm(ts_hydro_detrend ~ dummies_hydro)
summary(seas_means_model_hydro)
```

```
##
## Call:
## lm(formula = ts_hydro_detrend ~ dummies_hydro)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -33.933  -5.798  -0.531   5.721  32.166
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)        0.4379     1.4258   0.307 0.758849
## dummies_hydroJan   4.8863     2.0067   2.435 0.015177 *
## dummies_hydroFeb  -2.5567     2.0067  -1.274 0.203116
## dummies_hydroMar   7.0202     2.0067   3.498 0.000502 ***
## dummies_hydroApr   5.3770     2.0067   2.680 0.007572 **
## dummies_hydroMay  13.8957     2.0067   6.925 1.11e-11 ***
## dummies_hydroJun  10.7293     2.0067   5.347 1.27e-07 ***
## dummies_hydroJul   4.0439     2.0067   2.015 0.044320 *
## dummies_hydroAug  -5.3775     2.0067  -2.680 0.007566 **
```

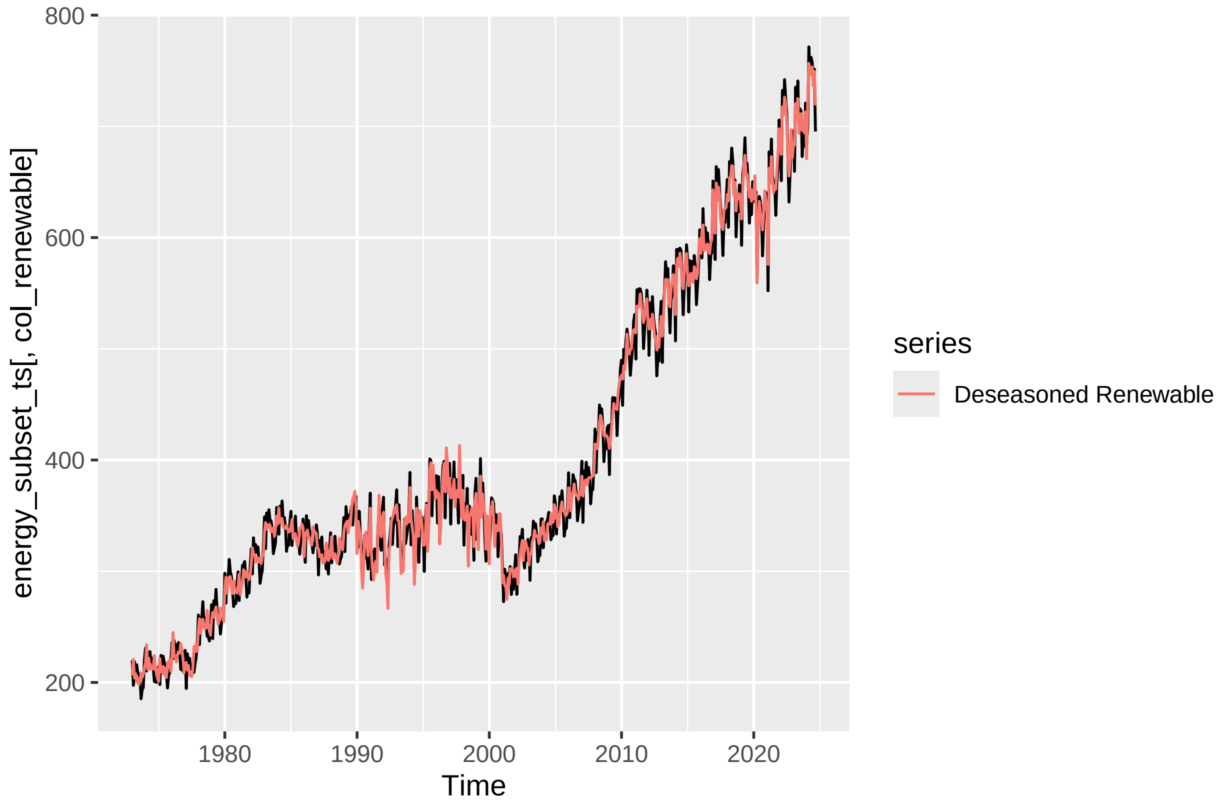
```
## dummies_hydroSep -16.5635     2.0067  -8.254 9.51e-16 ***
## dummies_hydroOct -16.3915     2.0164  -8.129 2.43e-15 ***
## dummies_hydroNov -10.8163     2.0164  -5.364 1.16e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.18 on 609 degrees of freedom
## Multiple R-squared:  0.4781, Adjusted R-squared:  0.4687
## F-statistic: 50.72 on 11 and 609 DF,  p-value: < 2.2e-16
```

```r
beta_intercept_hydro <- seas_means_model_hydro$coefficients[1]
beta_coeff_hydro <- seas_means_model_hydro$coefficients[2:12]
```

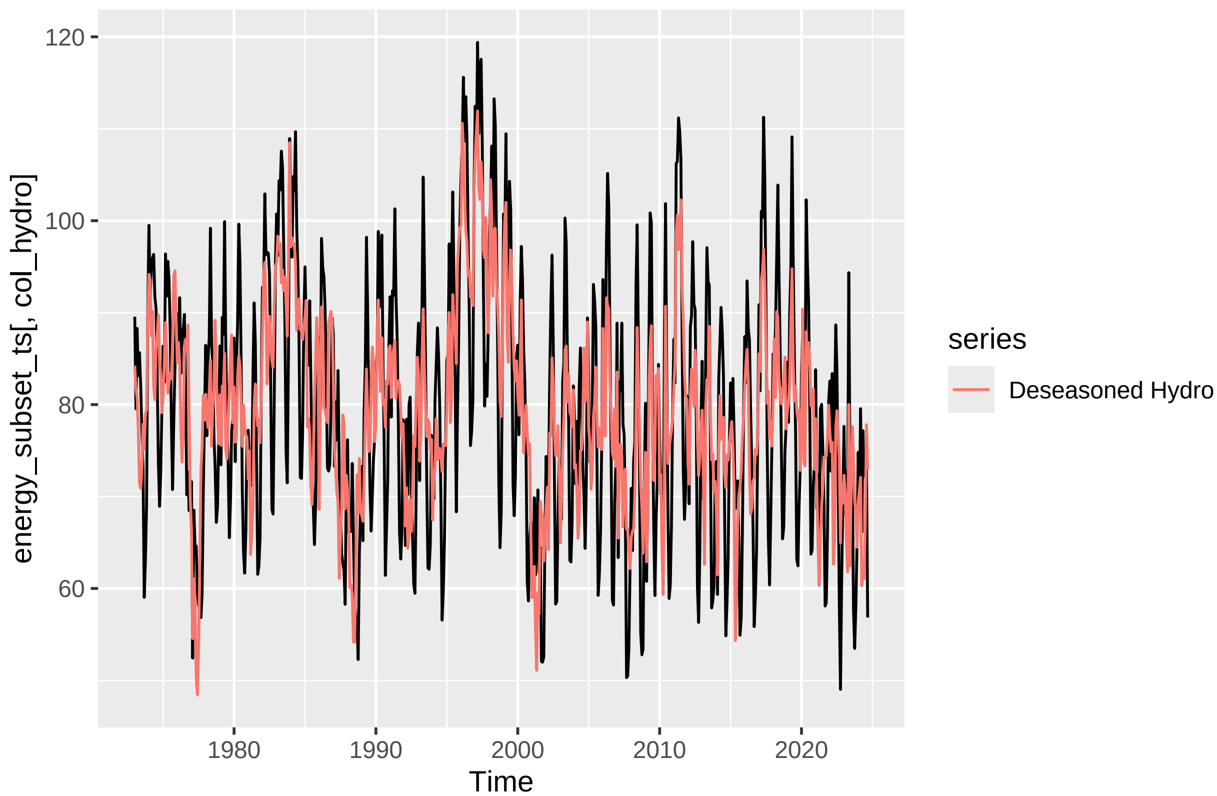
Q8

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
renewable_seas_comp <- array(0, nobs)
i <- 1
for(i in 1:nobs){
  renewable_seas_comp[i] <- beta_intercept_renewable + beta_coeff_renewable %*%
    dummies_renewable[i,]
}

deseason_renewable <- energy_subset[1] - renewable_seas_comp

ts_deseason_renewable_data <- ts(deseason_renewable, start = c(1973,1), frequency = 12)

autoplot(energy_subset_ts[,col_renewable])+
  autolayer(ts_deseason_renewable_data, series = "Deseasoned Renewable")
```



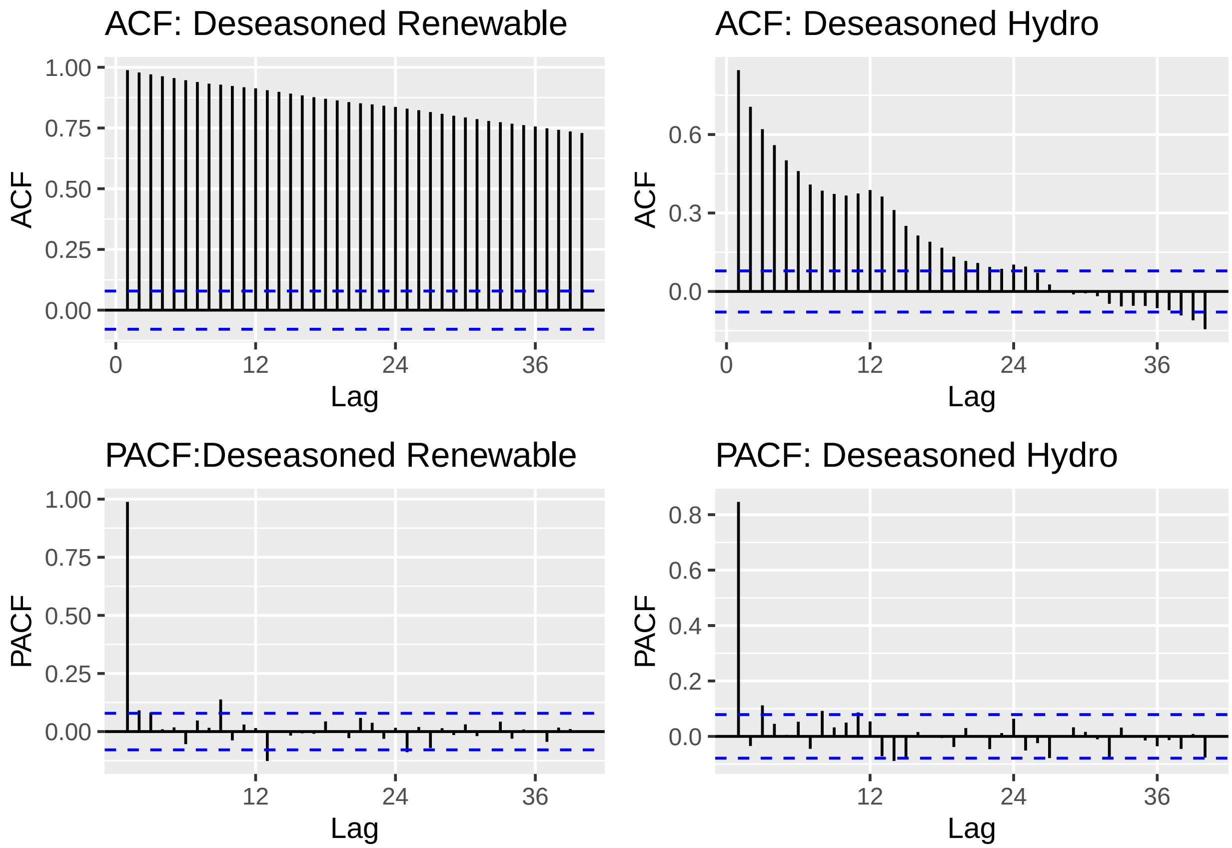

```
hydro_seas_comp <- array(0,nobs)
i <- 1
for(i in 1:nobs){
  hydro_seas_comp[i] <- beta_intercept_hydro + beta_coeff_hydro %*%
    dummies_hydro[i,]
}

deseason_hydro <- energy_subset[2] - hydro_seas_comp

ts_deseason_hydro_data <- ts(deseason_hydro, start = c(1973,1), frequency = 12)

autoplot(energy_subset_ts[,col_hydro])+
  autolayer(ts_deseason_hydro_data, series = "Deseasoned Hydro")
```



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, but not mandatory. Did the plots change? How?

```
# ACF and PACF using ggAcf() and ggPacf()
acf_renewable_deseason <- ggAcf(ts_deseason_renewable_data, lag.max = 40) +
  ggtitle(paste("ACF: Deseasoned Renewable"))
acf_hydro_deseason <- ggAcf(ts_deseason_hydro_data, lag.max = 40) +
  ggtitle(paste("ACF: Deseasoned Hydro"))

pacf_renewable_deseason <- ggPacf(ts_deseason_renewable_data, lag.max = 40) +
  ggtitle(paste("PACF:Deseasoned Renewable"))
pacf_hydro_deseason <- ggPacf(ts_deseason_hydro_data, lag.max = 40) +
  ggtitle(paste("PACF: Deseasoned Hydro"))

# Using plot_grid to arrange plots (all are now ggplot objects)
plot_grid(acf_renewable_deseason, acf_hydro_deseason, pacf_renewable_deseason,
          pacf_hydro_deseason, ncol = 2)
```


ACF: Deseasoned Renewable
ACF: Deseasoned Hydro
PACF:Deseasoned Renewable
PACF: Deseasoned Hydro
ACF
PACF
Lag
1.00
0.75
0.50
0.25
0.00
0.6
0.3
0.0
0.8
0.4
0.2
0
12
24
36