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

Assignment 3 - Due date 02/04/25

Aye Nyein Thu

## 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.\

```r
# Load/install required package here
library(readxl)
library(lubridate)
library(dplyr)
library(cowplot)
library(forecast)
library(ggplot2)
library(Kendall)
library(tseries)

# Set my theme
mytheme <- theme_classic() +
```

```r
  theme(axis.text = element_text(color = "black"),
        legend.position = "top",
        plot.title = element_text(hjust = 0.5, face = "bold"))

theme_set(mytheme)
```

```r
# Import Dataset
Energy <- read_excel(
  path="./Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source.xlsx",
  skip = 12, sheet="Monthly Data",col_names=FALSE)

# Extract the column names from row 11
Column_Names <- read_excel(
  path="./Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source.xlsx",
  skip = 10,n_max = 1, sheet="Monthly Data", col_names=FALSE)

# Assign the column names to the data set
colnames(Energy) <- Column_Names

# Select the columns of interest and format date
Energy_cleaned <- Energy %>%
  select("Month",
         "Total Renewable Energy Production",
         "Hydroelectric Power Consumption") %>%
  mutate(Month=as.Date(Month, format="%Y-%m-%d")) %>%
  rename(Renewable = "Total Renewable Energy Production",
         Hydroelectric = "Hydroelectric Power Consumption")

# Assign values to no. of observation
nobs <- nrow(Energy_cleaned)
```

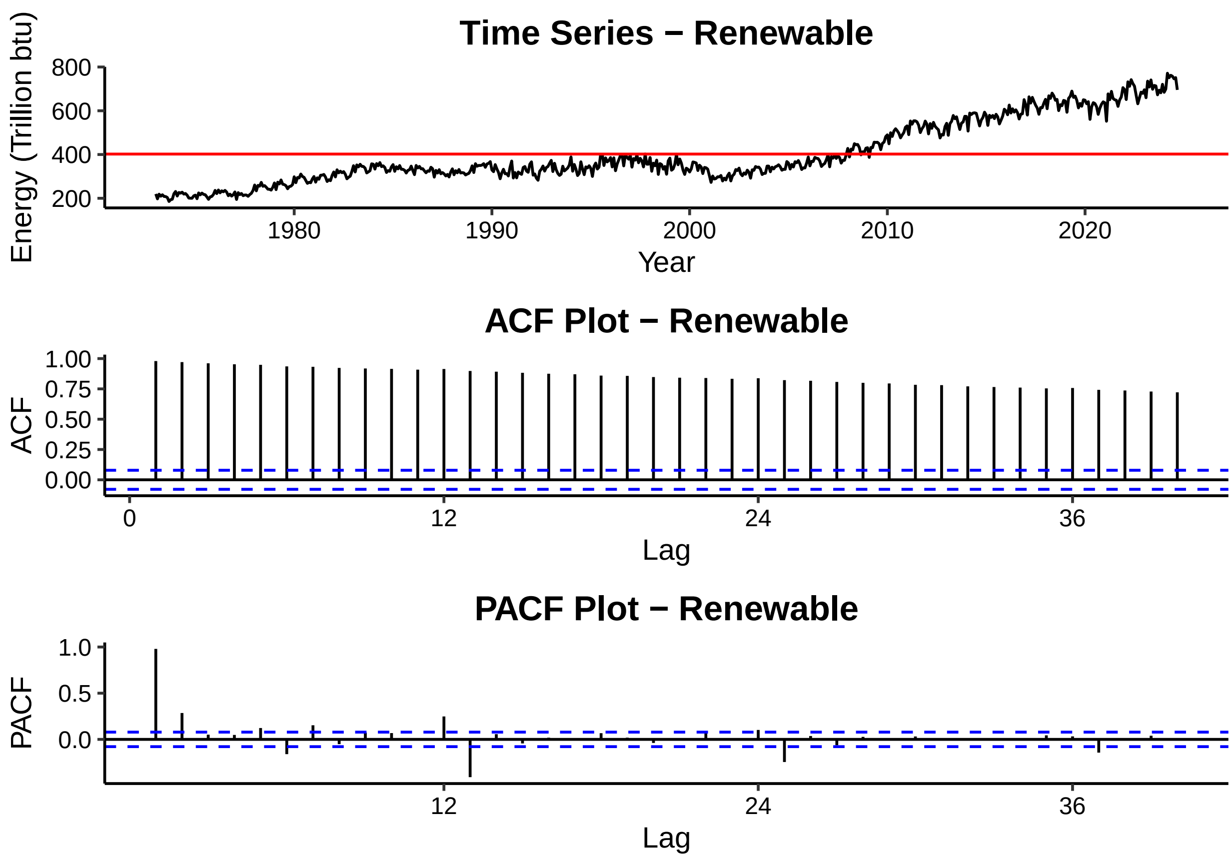
## Trend Component

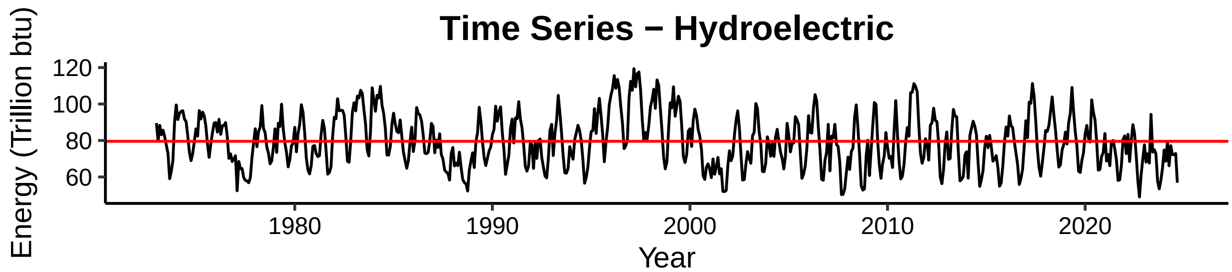
### Q1

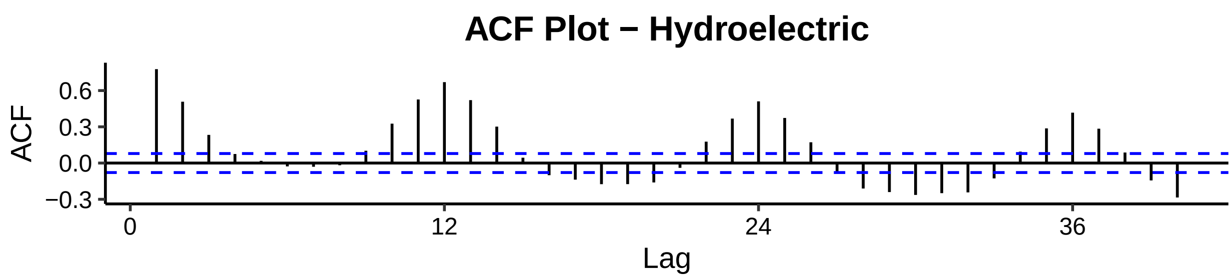
For each time series, i.e., Renewable Energy Production and Hydroelectric Consumption create three plots: one with time series, one with the ACF and with the PACF. You may use the some code from A2, but I want all the three plots side by side as in a grid. (Hint: use function `plot_grid()` from the `cowplot` package)

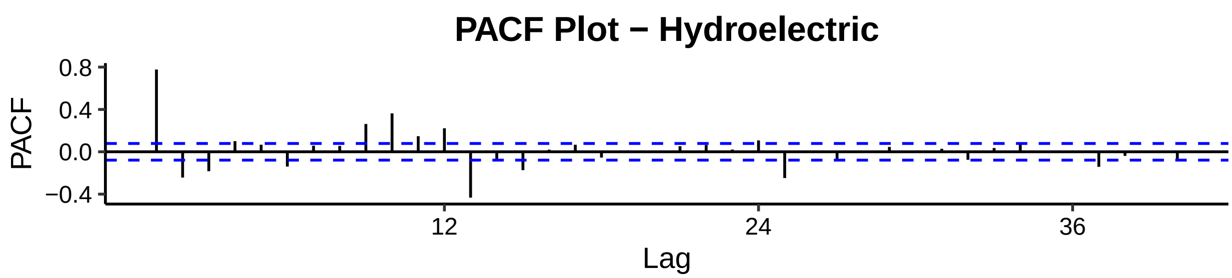
```r
# Specify starting points
start_year <- as.numeric(format(min(Energy_cleaned$Month), "%Y"))
start_month <- as.numeric(format(min(Energy_cleaned$Month), "%m"))

# Transform data frame to time series objects
Energy_ts <- ts(Energy_cleaned[,2:3],
                start = c(start_year, start_month), frequency = 12)

# Create plots for time series, ACF and PACF for two series
for(i in 1:2) {

  # Calculate mean of each time series
  avg_value <- mean(Energy_ts[, i])
```

```r
  # Plot time series
  ts_plot <- autoplot(Energy_ts[, i],
                      xlab = "Year", ylab = "Energy (Trillion btu)") +
             geom_hline(yintercept = avg_value, color = "red") +
             ggtitle(paste("Time Series -", colnames(Energy_ts)[i]))

  # Plot ACF
  acf_plot <- ggAcf(Energy_ts[, i], lag.max = 40) +
    ggtitle(paste("ACF Plot -", colnames(Energy_ts)[i]))

  # Plot PACF
  pacf_plot <- ggPacf(Energy_ts[, i], lag.max = 40) +
    ggtitle(paste("PACF Plot -", colnames(Energy_ts)[i]))

  # Combine plots
  combined_plot <- plot_grid(ts_plot, acf_plot, pacf_plot, nrow = 3,
                             align = "v")

  # Print the combined plots for each time series
  print(combined_plot)
}
```






**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?

> As depicted in Q1, the total renewable energy production has an increasing trend. Despite a slight reduction in production volume during 2000 and 2008, it generally exhibits an upward trend. The autocorelation between the time points are also strong.
>
> On the contrary, the hydroelectric power consumption does not have an apparent increasing or decreasing trend. The autocorrelation plot also shows that the spikes at 12, 24 and 36 lags and the mix of both positive and negative values indicates a seasonal structure rather than an upward or downward trend.
>
> The partial autocorrelation plots for both series indicates the significant spikes in 13, 25 and 37 lags which could be seasonal behaviours.

**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
# Create vector t
t <- c(1:nobs)

# Fit a linear model for Renewable
linear_renew <- lm(Energy_cleaned$Renewable ~ t)
summary(linear_renew)
```

```
##
## Call:
## lm(formula = Energy_cleaned$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
# Save the regression coefficients for Renewable
beta0_linear_renew <- as.numeric(linear_renew$coefficients[1])
beta1_linear_renew <- as.numeric(linear_renew$coefficients[2])

# Fit a linear model for Hydroelectric
linear_hydro <- lm(Energy_cleaned$Hydroelectric ~ t)
summary(linear_hydro)
```

```
##
## Call:
## lm(formula = Energy_cleaned$Hydroelectric ~ 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
```

```r
# Save the regression coefficients for Hydroelectric
beta0_linear_hydro <- as.numeric(linear_hydro$coefficients[1])
beta1_linear_hydro <- as.numeric(linear_hydro$coefficients[2])
```

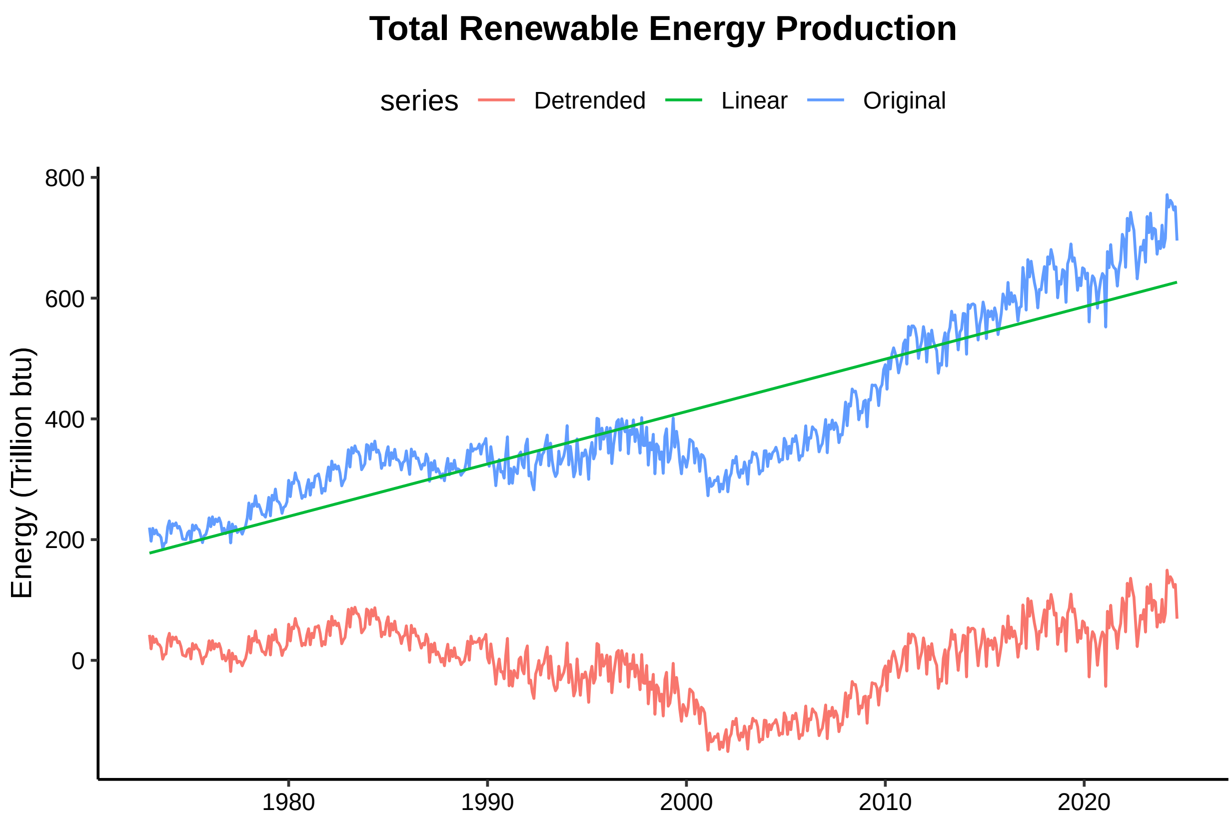
> The linear model for the total renewable production indicates that the time vector (t) exhibits a positive coefficient and is statistically significant at 99% level. Therefore, a unit increase in time is associated with an increase in total renewable energy production. The intercept value is 176.87 trillion btu if time (t) is zero.

> When the total hydroelectric power consumption is fitted in the linear model, it exhibits a significant negative coefficient although the magnitude of the coefficient is small compared to the one for renewable. Although the plot in Question 1 does not exhibit a strong decreasing trend, the significant and negative coefficient in the linear model indicates that an increase in time unit is associated with the decrease in total hydroelectric consumption in general.

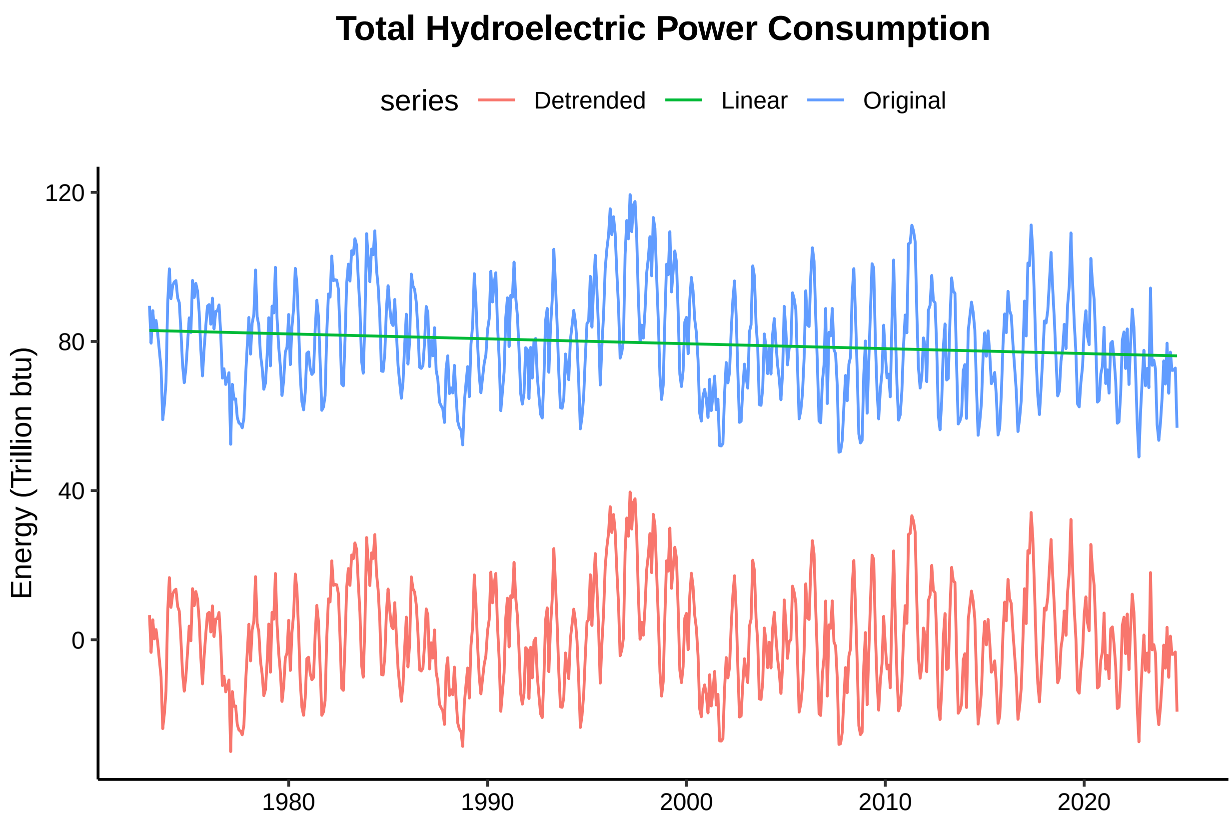
### Q4

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
# Create detrended series from the linear model for renewable
linear_trend_renew <- beta0_linear_renew + beta1_linear_renew * t
ts_linear_trend_renew <- ts(linear_trend_renew,
                            start=c(start_year,start_month), frequency=12)

detrend_renew <- Energy_cleaned[,2] - linear_trend_renew
ts_detrend_renew <- ts(detrend_renew,
                       start=c(start_year,start_month), frequency = 12)

# Plot the detrended series for renewable
autoplot(Energy_ts[,1], series = "Original") +
  autolayer(ts_detrend_renew, series = "Detrended") +
  autolayer(ts_linear_trend_renew, series = "Linear") +
  labs(title = "Total Renewable Energy Production",
       x = "", y = "Energy (Trillion btu)")
```




```
# Create detrended series from the linear model for hydroelectric
linear_trend_hydro <- beta0_linear_hydro + beta1_linear_hydro * t
ts_linear_trend_hydro <- ts(linear_trend_hydro,
                            start=c(start_year,start_month), frequency=12)

detrend_hydro <- Energy_cleaned[,3] - linear_trend_hydro
ts_detrend_hydro <- ts(detrend_hydro,
                       start=c(start_year,start_month), frequency = 12)

# Plot the detrended series for hydroelectric
autoplot(Energy_ts[,2], series = "Original") +
  autolayer(ts_detrend_hydro, series = "Detrended") +
  autolayer(ts_linear_trend_hydro, series = "Linear") +
  labs(title = "Total Hydroelectric Power Consumption",
       x = "", y = "Energy (Trillion btu)")
```




For total renewable energy production, the original time series plot in Question 1 exhibits a strong increasing trend. When the upward trend component is removed, the detrended series do not show an apparent upward trend and there was a low production during the period between 2003 and 2010.

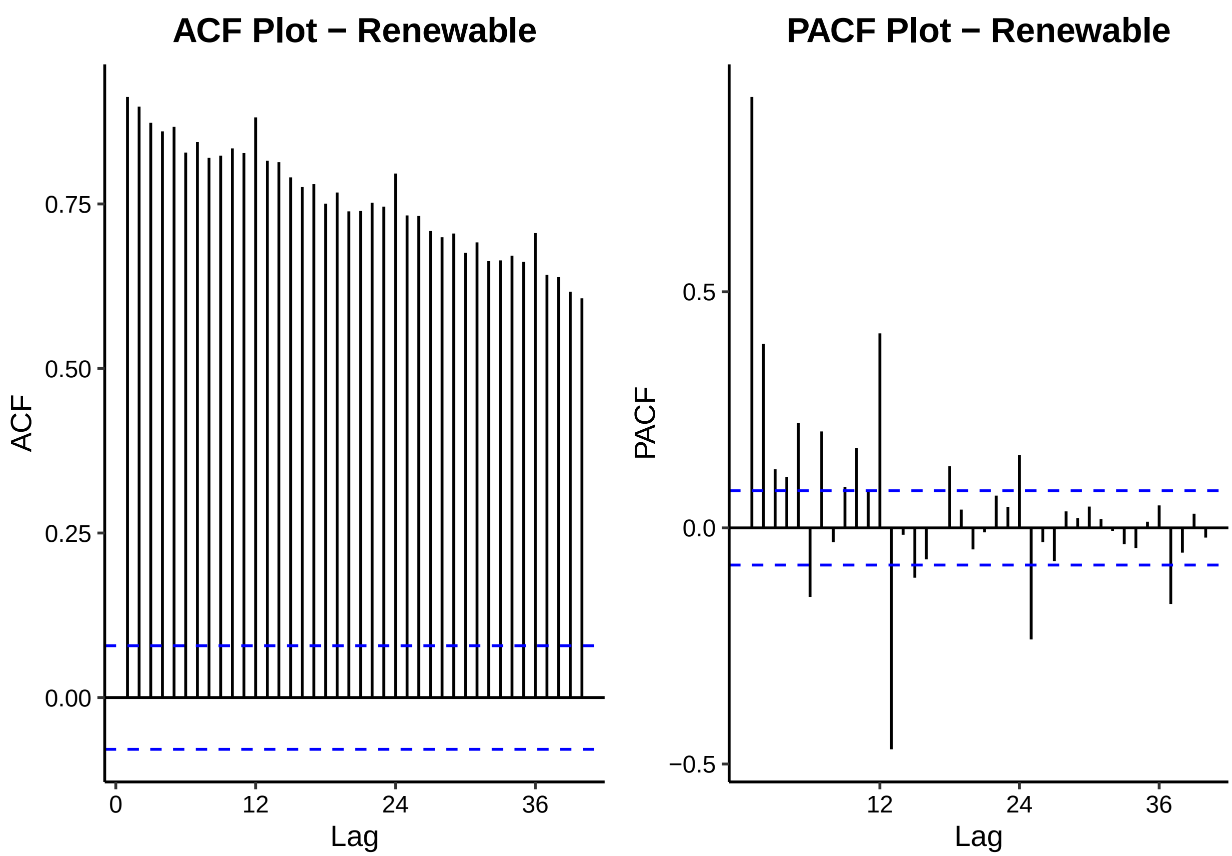
For total hydroelectric power consumption, the original time series plot in Question 1 do not show any increasing or decreasing trend. Therefore, the linear trend line in Question 5 plot indicates a weak downward-leaning line. Compared to the plot in Question 1, the detrended series in Question 5 surely decreases the values, but they are more or less in similar feature with no apparent trend line. The linear trend line might be different if the moving average method is used instead of linear model.

**Q5**

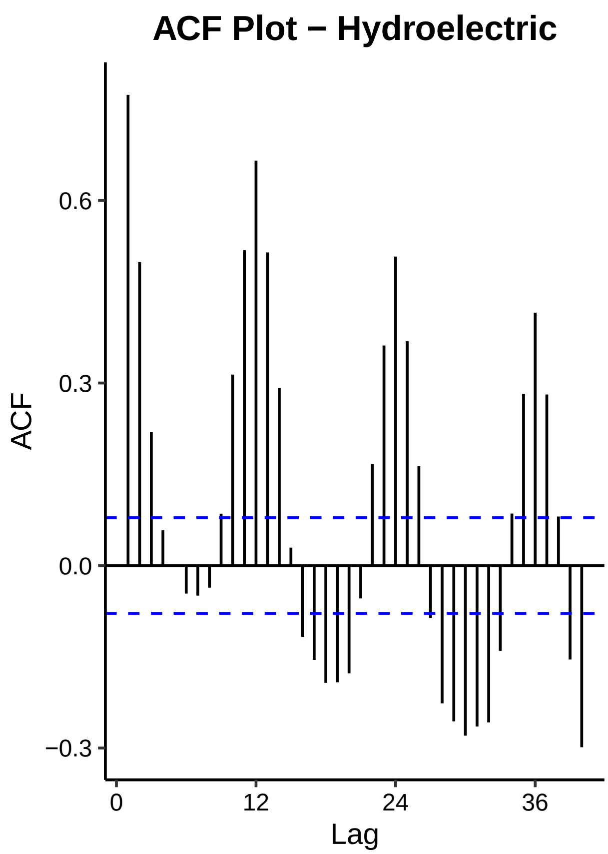
Plot ACF and PACF for the detrended series and compare with the plots from Q1. You may use plot_grid() again to get them side by side, but not mandatory. Did the plots change? How?

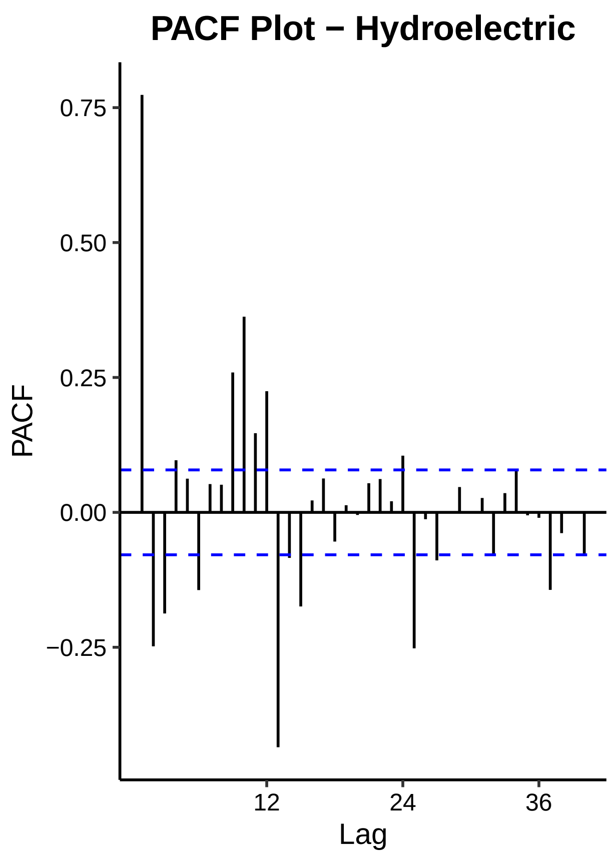
```
# Plot ACF for renewable
acf_detrend_renew <- ggAcf(ts_detrend_renew, lag.max = 40) +
  ggtitle(paste("ACF Plot - Renewable"))

# Plot PACF for renewable
pacf_detrend_renew <- ggPacf(ts_detrend_renew, lag.max = 40) +
    ggtitle(paste("PACF Plot - Renewable"))
```

```r
# Combine plots for renewable
plot_grid(acf_detrend_renew, pacf_detrend_renew, ncol = 2)
```




```r
# Plot ACF for hydroelectric
acf_detrend_hydro <- ggAcf(ts_detrend_hydro, lag.max = 40) +
  ggtitle(paste("ACF Plot - Hydroelectric"))

# Plot PACF for hydroelectric
pacf_detrend_hydro <- ggPacf(ts_detrend_hydro, lag.max = 40) +
    ggtitle(paste("PACF Plot - Hydroelectric"))

# Combine plots for hydroelectric
plot_grid(acf_detrend_hydro, pacf_detrend_hydro, ncol = 2)
```




The ACF plot in Question 1 for renewable production shows a strong dependency between time values. Although the element of strong autocorrelation is still there in the detrended series, they are no longer consistently on the upward trend. The PACF plots in both Question 1 and 5 highlights the possibility of seasonality due to spikes in lags at 13, 25 and 37 respectively.

The ACF and PACF plots in Question 1 and 5 for hydroelectric power consumption do not have much differences. It is explained by the series' nature that does not exhibit a significant trend. As in the Qusestion 1, the spikes in PACF plot at 13, 25 and 37 have a high chance of seasonality. However, it is also necessary to conduct a more rigorous analysis to prove the conclusion.

## 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 renewable production, just the detrended series plot alone makes it difficult to conclude that there is seasonality behavior in the data. However, the spikes in PACF plots shows the possibility of seasonality. Therefore, it requires further analysis on deseasoning to make a valid conclusion.

However, hydroelectric power consumption indicates the different narration. Both original time series and detrended time series do not show a strong trend. Despite, the spikes in both ACF and PACF exhibits a possibility of seasonality.

### 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?

```r
# Use seasonal means model for renewable
dummies_renew <- seasonaldummy(ts_detrend_renew)

seas_means_renew <- lm(detrend_renew$Renewable ~ dummies_renew)
summary(seas_means_renew)
```

```
##
## Call:
## lm(formula = detrend_renew$Renewable ~ dummies_renew)
##
## 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_renewJan    5.592     11.968   0.467  0.64048
## dummies_renewFeb  -31.452     11.968  -2.628  0.00881 **
## dummies_renewMar    6.892     11.968   0.576  0.56491
## dummies_renewApr   -6.449     11.968  -0.539  0.59023
## dummies_renewMay    7.923     11.968   0.662  0.50822
## dummies_renewJun   -3.394     11.968  -0.284  0.77682
## dummies_renewJul    2.126     11.968   0.178  0.85906
## dummies_renewAug   -5.878     11.968  -0.491  0.62351
## dummies_renewSep  -31.209     11.968  -2.608  0.00934 **
## dummies_renewOct  -18.757     12.026  -1.560  0.11937
## dummies_renewNov  -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
```

```r
beta0_seas_renew <-seas_means_renew $coefficients[1]
beta1_seas_renew <-seas_means_renew$coefficients[2:12]

# Use seasonal means model for hydroelectric
dummies_hydro <- seasonaldummy(ts_detrend_hydro)

seas_means_hydro <- lm(detrend_hydro$Hydroelectric ~ dummies_hydro)
summary(seas_means_hydro)
```

```
##
## Call:
## lm(formula = detrend_hydro$Hydroelectric ~ 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
beta0_seas_hydro <-seas_means_hydro $coefficients[1]
beta1_seas_hydro <-seas_means_hydro$coefficients[2:12]
```

The regression outputs in Question 7 are in line with the estimation made in Question 6. As expected, seasonality is more pronounced in the hydroelectric series with majority of months as dummies have strong and significant coefficients. While March to July show positive coefficients and September to November have negative ones.

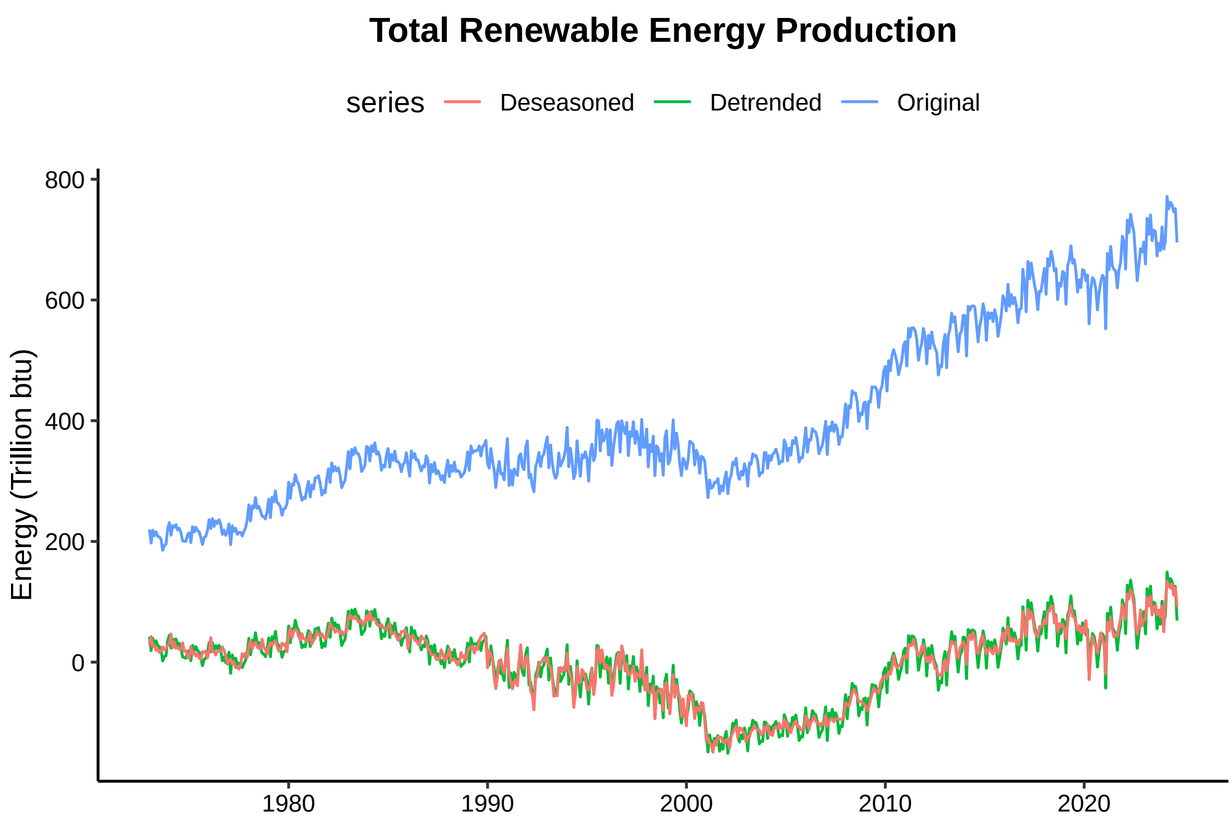
However, in the case of renewable production, only February and September have significant negative coefficients while the rest are not statistically significant, leading it difficult to conclude the existence of seasonality.

### 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?

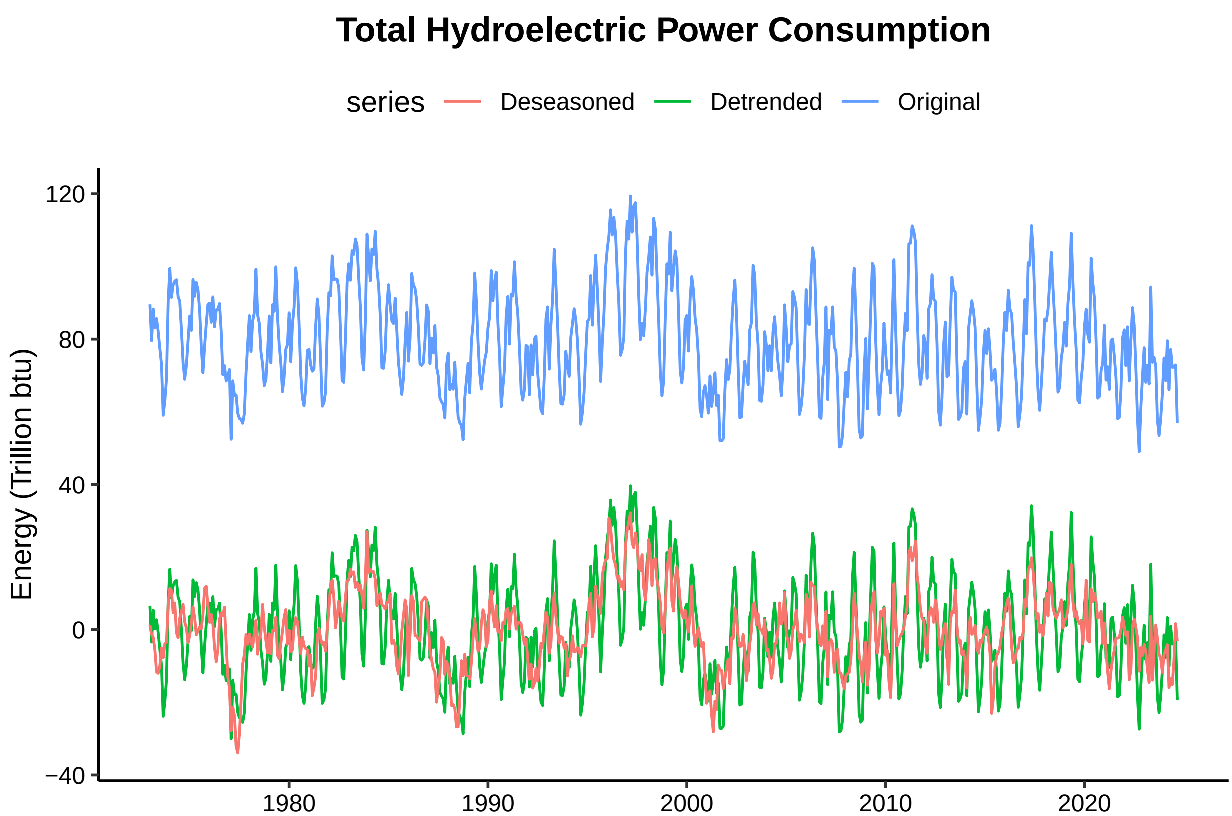
```r
# Deseason renewable
seas_renew <- array(0,nobs)

for(i in 1:nobs){
  seas_renew[i] <- beta0_seas_renew + beta1_seas_renew %*% dummies_renew[i,]
}
```

```
deseason_renew <- detrend_renew - seas_renew

ts_deseason_renew <- ts(deseason_renew,
                        start=c(start_year,start_month), frequency = 12)

# Plot the detrended series for renewable
autoplot(Energy_ts[,1], series = "Original") +
  autolayer(ts_detrend_renew, series = "Detrended") +
  autolayer(ts_deseason_renew, series = "Deseasoned") +
  labs(title = "Total Renewable Energy Production",
       x = "", y = "Energy (Trillion btu)")
```




```
# Deseason hydroelectric
seas_hydro <- array(0,nobs)

for(i in 1:nobs){
  seas_hydro[i] <- beta0_seas_hydro + beta1_seas_hydro %*% dummies_hydro[i,]
}

deseason_hydro <- detrend_hydro - seas_hydro

ts_deseason_hydro <- ts(deseason_hydro,
                        start=c(start_year,start_month), frequency = 12)

# Plot the detrended series for renewable
```

```r
autoplot(Energy_ts[,2], series = "Original") +
  autolayer(ts_detrend_hydro, series = "Detrended") +
  autolayer(ts_deseason_hydro, series = "Deseasoned") +
  labs(title = "Total Hydroelectric Power Consumption",
       x = "", y = "Energy (Trillion btu)")
```



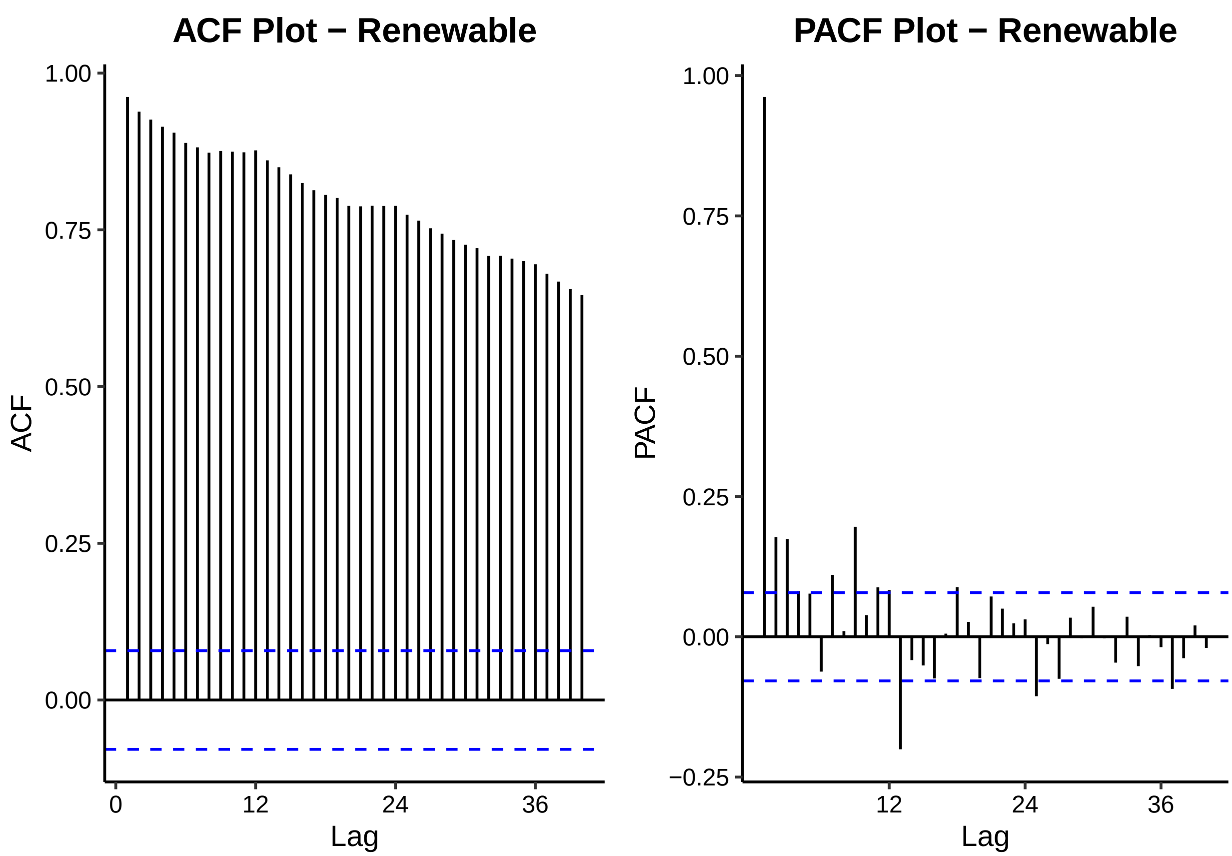

Compared to Question 1, the plots in Question 8 for both renewable production and hydroelectric power consumption show a significant change. Before introducing detrending and deseasoning, the renewable production series exhibits a strong upward trend while the hydroelectric power is more in seasonal nature. After deseasoning the series, both trends and seasons elements have flatted out, the series are now in the smaller data ranges and more likely to be in random nature.

### Q9

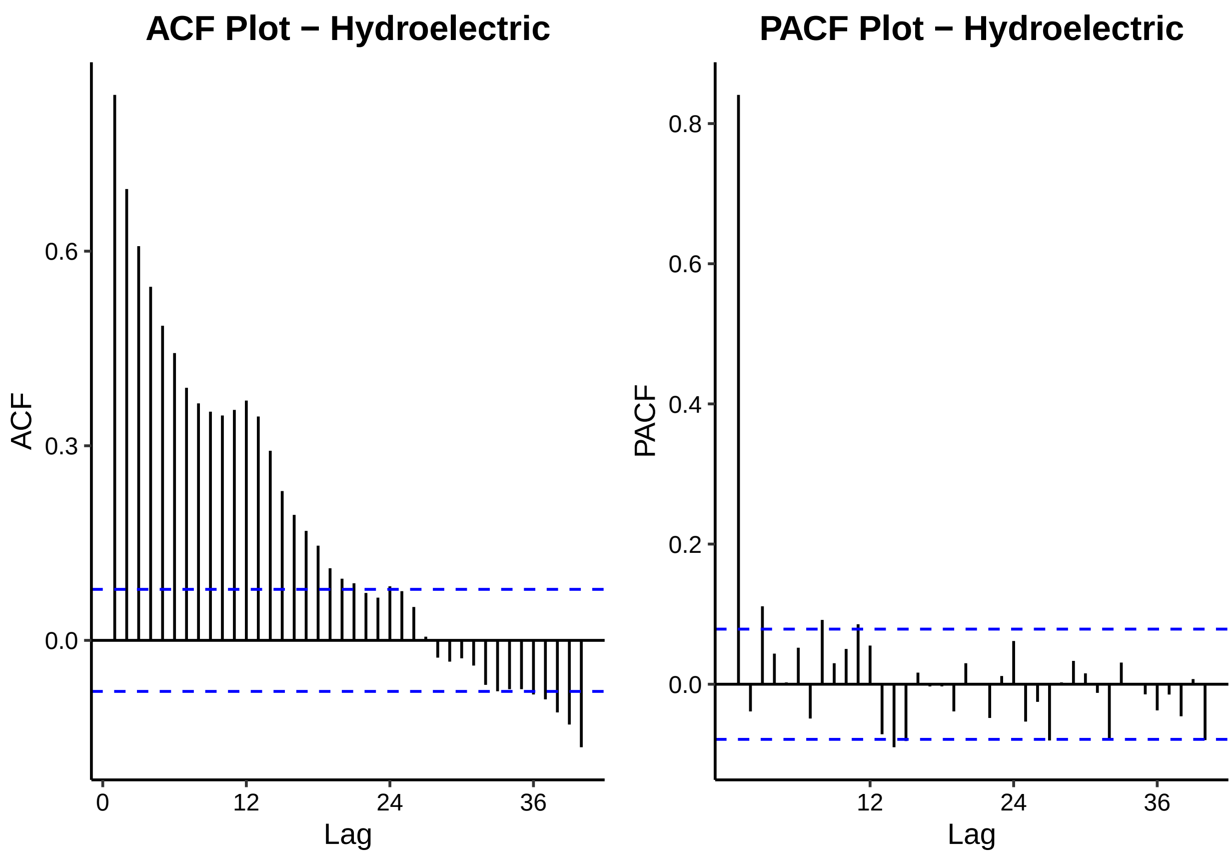
Plot ACF and PACF for the deseason series and compare with the plots from Q1. You may use plot_grid() again to get them side by side, but not mandatory. Did the plots change? How?

```r
# Plot ACF for renewable
acf_deseason_renew <- ggAcf(ts_deseason_renew, lag.max = 40) +
  ggtitle(paste("ACF Plot - Renewable"))

# Plot PACF for renewable
pacf_deseason_renew <- ggPacf(ts_deseason_renew, lag.max = 40) +
    ggtitle(paste("PACF Plot - Renewable"))
```

```r
# Combine plots for renewable
plot_grid(acf_deseason_renew, pacf_deseason_renew, ncol = 2)
```




```r
# Plot ACF for hydroelectric
acf_deseason_hydro <- ggAcf(ts_deseason_hydro, lag.max = 40) +
  ggtitle(paste("ACF Plot - Hydroelectric"))

# Plot PACF for hydroelectric
pacf_deseason_hydro <- ggPacf(ts_deseason_hydro, lag.max = 40) +
    ggtitle(paste("PACF Plot - Hydroelectric"))

# Combine plots for hydroelectric
plot_grid(acf_deseason_hydro, pacf_deseason_hydro, ncol = 2)
```




After deseasoning the series, the renewable energy production series still exhibit a strong autocorrelation between time points, however, the seasonality possibility is less pronounced. For the hydroelectric series, their seasonality effects are more flattened out in the deseasoned PACF plot.