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

Assignment 3 - Due date 02/04/25

Yuqi Yang

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(openxlsx)
library(dplyr)
library(forecast)
library(tseries)
library(Kendall)
library(cowplot)
library(ggplot2)
```

```
#Importing data set
raw_energy_data <- read.xlsx(xlsxFile = "../Data/Table_10.1_Renewable_Energy_Production_and_Consumption_
```

```r
                              sheet = "Monthly Data",
                              startRow = 13,
                              colNames = FALSE,
                              detectDates=TRUE)

read_col_names <- read.xlsx(xlsxFile = "../Data/Table_10.1_Renewable_Energy_Production_and_Consumption_
                             sheet = "Monthly Data",
                             rows = 11,
                             colNames = FALSE,
                             detectDates=TRUE)

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

#Selecting the columns of interest
energy_data <- raw_energy_data %>%
select("Month",
"Total Renewable Energy Production",
"Hydroelectric Power Consumption") %>%
mutate(Month = as.Date(Month, origin = "1899-12-30", format = "%Y/%m/%d"))

#Verifying data
head(energy_data)
```

```
##        Month Total Renewable Energy Production Hydroelectric Power Consumption
## 1 1973-01-01                           219.839                          89.562
## 2 1973-02-01                           197.330                          79.544
## 3 1973-03-01                           218.686                          88.284
## 4 1973-04-01                           209.330                          83.152
## 5 1973-05-01                           215.982                          85.643
## 6 1973-06-01                           208.249                          82.060
```

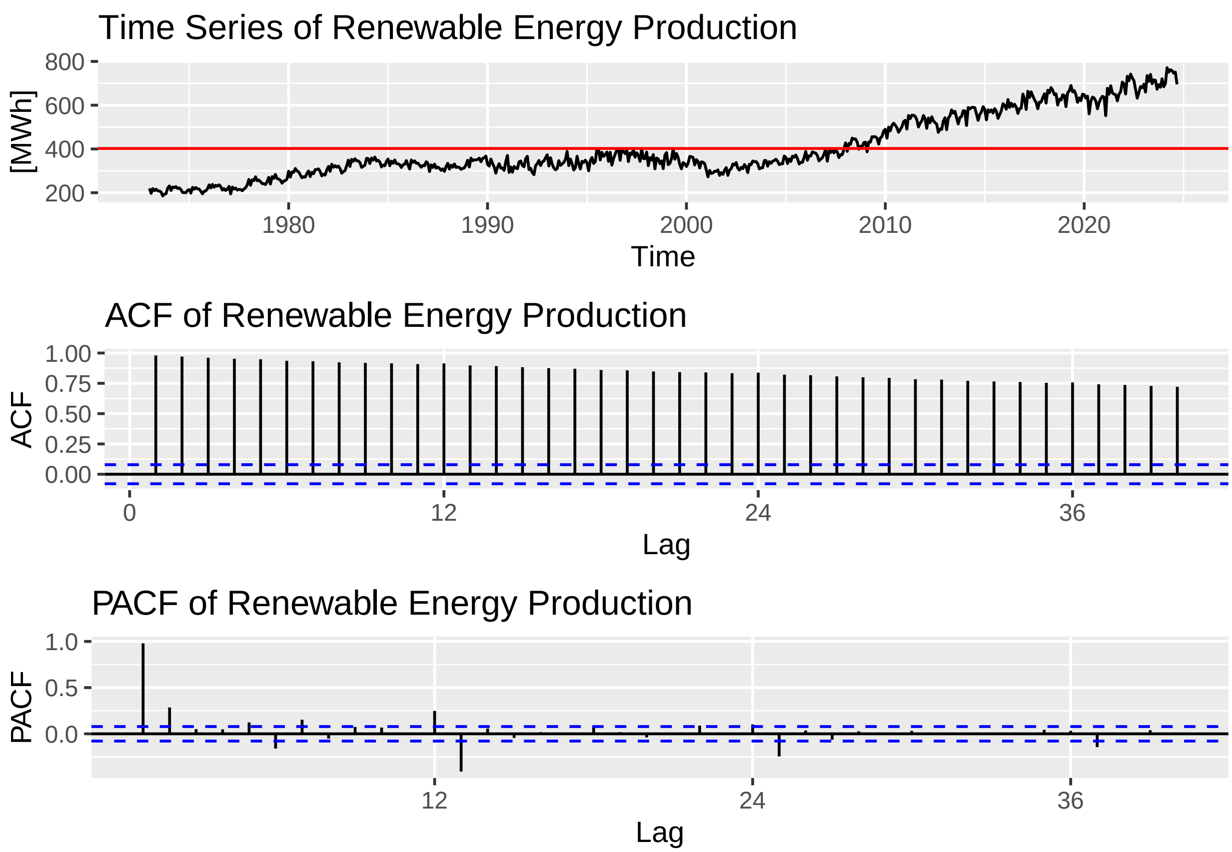
##Trend Component

Q1

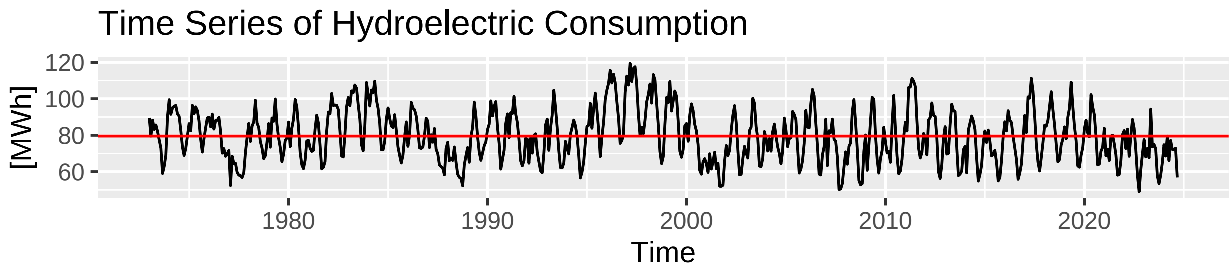
For each time series, i.e., Renewable Energy Production and Hydroelectric Consumption create three plots: one with time series, one with the ACF and with the PACF. You may use the some code form A2, but I want all the three plots side by side as in a grid. (Hint: use function `plot_grid()` from the `cowplot`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

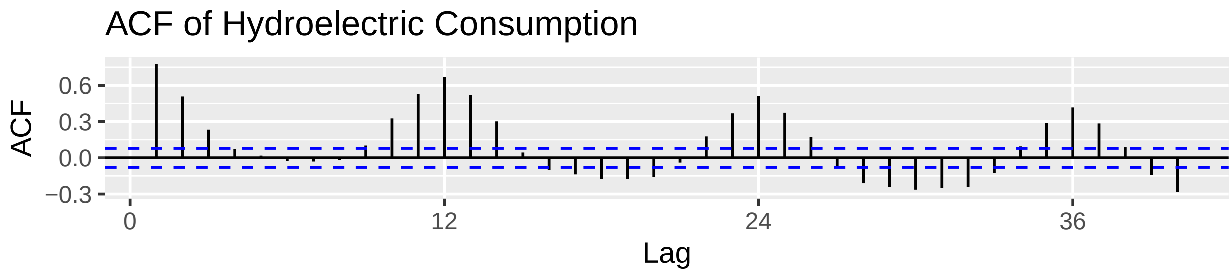
```r
#Transforming data into time series object
ts_energy_data <- ts(energy_data[,2:3], start=c(1973,1), frequency=12)

#Function to create plots
plot_series <- function(ts_data, title) {
  ts_plot <- autoplot(ts_data) + ggtitle(paste("Time Series of", title)) +
    geom_hline(yintercept = mean(ts_data),color = "red") +
    ylab("[MWh]")
  acf_plot <- ggAcf(ts_data, lag.max=40) + ggtitle(paste("ACF of", title))
  pacf_plot <- ggPacf(ts_data, lag.max=40) + ggtitle(paste("PACF of", title))

  plot_grid(ts_plot, acf_plot, pacf_plot, nrow=3)
```

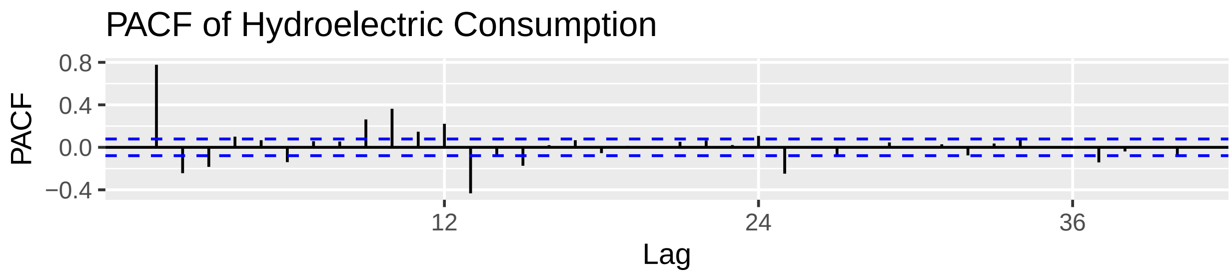
```r
}

#Plots for each time series
renewable_plot <- plot_series(ts_energy_data[,1], "Renewable Energy Production")
hydro_plot <- plot_series(ts_energy_data[,2], "Hydroelectric Consumption")

print(renewable_plot)
```




```r
print(hydro_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?

For **Total Renewable Energy Production**, it shows a strong upward trend, but this growth is non-linear. While **Hydroelectric Power Consumption** does not show a significant trend. It fluctuates around a relatively stable mean, but with some seasonal variation.

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
nobs <- nrow(energy_data)
t <- c(1:nobs)

#Fit a linear trend to the two time series
linear_trend_model_renewable <- lm(energy_data$`Total Renewable Energy Production`~ t)
linear_trend_model_hydro <- lm(energy_data$`Hydroelectric Power Consumption` ~ t)

summary(linear_trend_model_renewable)
```

```
##
## Call:
## lm(formula = energy_data$`Total Renewable Energy Production` ~
##     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
summary(linear_trend_model_hydro)
```

```
##
## Call:
## lm(formula = energy_data$`Hydroelectric Power Consumption` ~
##     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
```

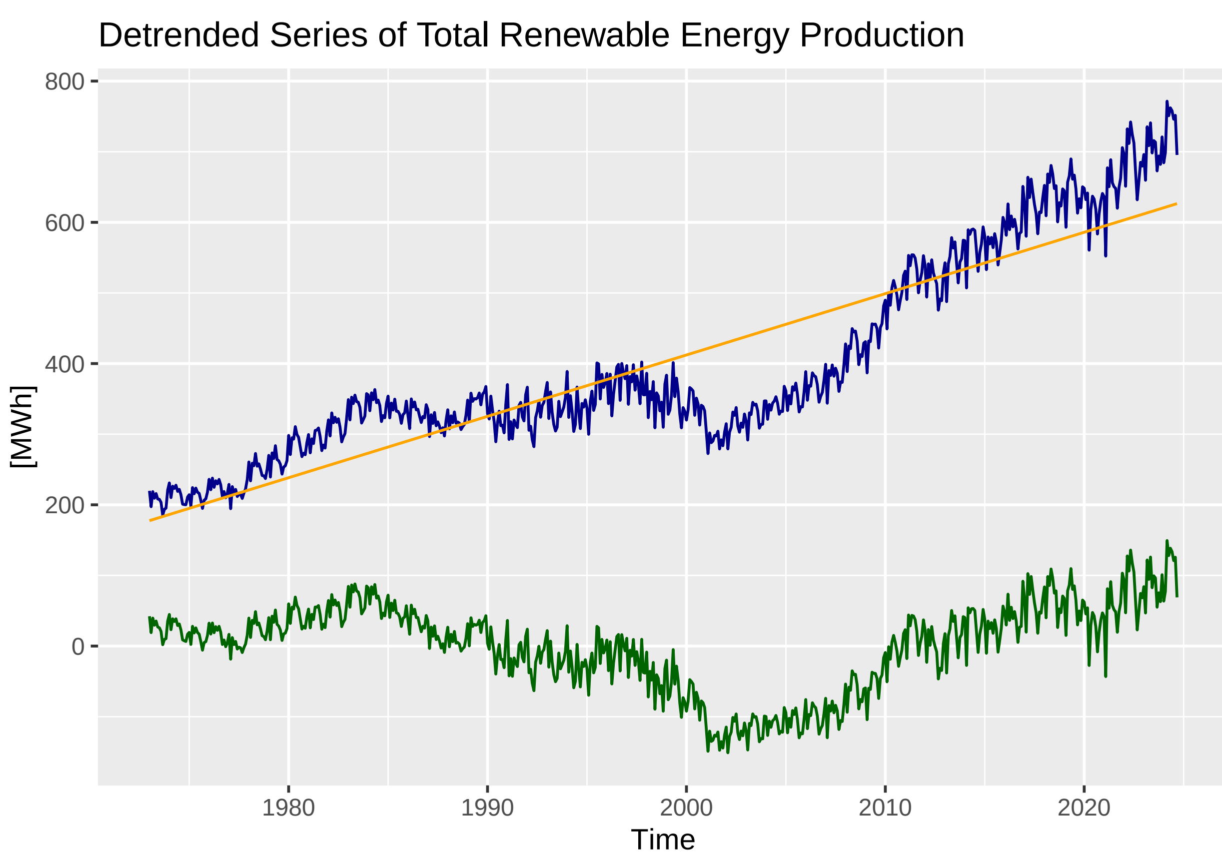
```r
#Save coefficients for further analysis
beta0_renewable <- coef(linear_trend_model_renewable)[1]
beta1_renewable <- coef(linear_trend_model_renewable)[2]

beta0_hydro <- coef(linear_trend_model_hydro)[1]
beta1_hydro <- coef(linear_trend_model_hydro)[2]
```

For **Renewable Energy Production**, the intercept is 176.87293, the slope is 0.72393, it is a strong upward trend.
For **Hydroelectric Consumption**, the intercept is 82.96766, the slope is -0.01098, it is a slight downward trend.

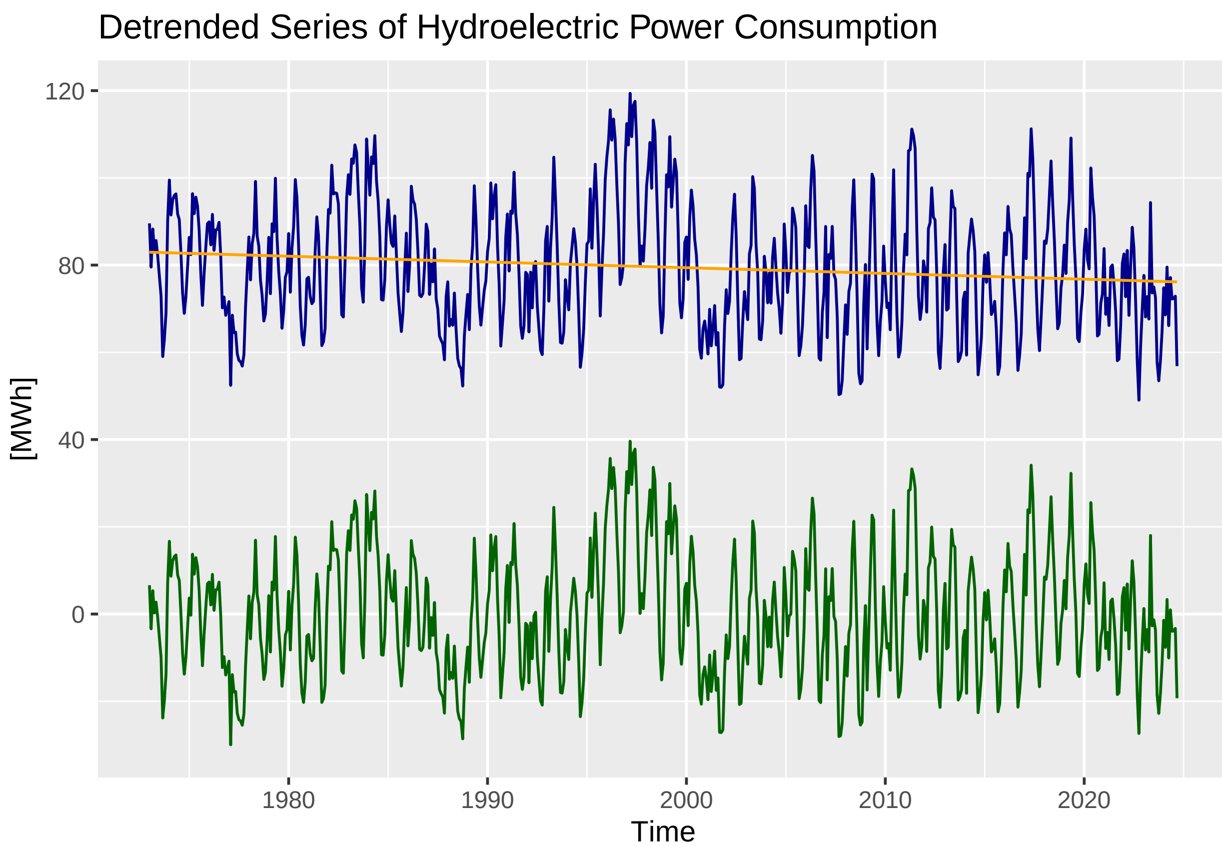
Q4

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
#Detrend the series
linear_trend_renewable <- beta0_renewable + beta1_renewable * t
linear_trend_hydro <- beta0_hydro + beta1_hydro * t

detrend_renewable <- energy_data$`Total Renewable Energy Production` - linear_trend_renewable
detrend_hydro <- energy_data$`Hydroelectric Power Consumption` - linear_trend_hydro

#Transform into time series objects
ts_linear_renewable <- ts(linear_trend_renewable, start = c(1973,1), frequency = 12)
ts_linear_hydro <- ts(linear_trend_hydro, start = c(1973,1), frequency = 12)
ts_detrend_renewable <- ts(detrend_renewable, start = c(1973,1), frequency = 12)
ts_detrend_hydro <- ts(detrend_hydro, start = c(1973,1), frequency = 12)

#Plot the detrended series for Renewable Energy Production
autoplot(ts_energy_data[,1], color="darkblue") +
  autolayer(ts_detrend_renewable, series="Detrended", color="darkgreen") +
  autolayer(ts_linear_renewable, series="Linear Component", color="orange") +
  ggtitle("Detrended Series of Total Renewable Energy Production") +
  ylab("[MWh]")
```



```r
#Plot the detrended series for Hydroelectric Consumption
autoplot(ts_energy_data[,2], color="darkblue") +
  autolayer(ts_detrend_hydro, series="Detrended", color="darkgreen") +
  autolayer(ts_linear_hydro, series="Linear Component", color="orange") +
  ggtitle("Detrended Series of Hydroelectric Power Consumption") +
  ylab("[MWh]")
```



For **Detrended Total Renewable Energy Production**, unlike the long-term upward trend shown in the Q1, now shows periodic fluctuations, indicating a possible seasonal pattern. And it has some upward trend after 2000, except for a short decline around 2020 due to the pandemic, possibly due to technological advances, policy adjustments, or other reasons.

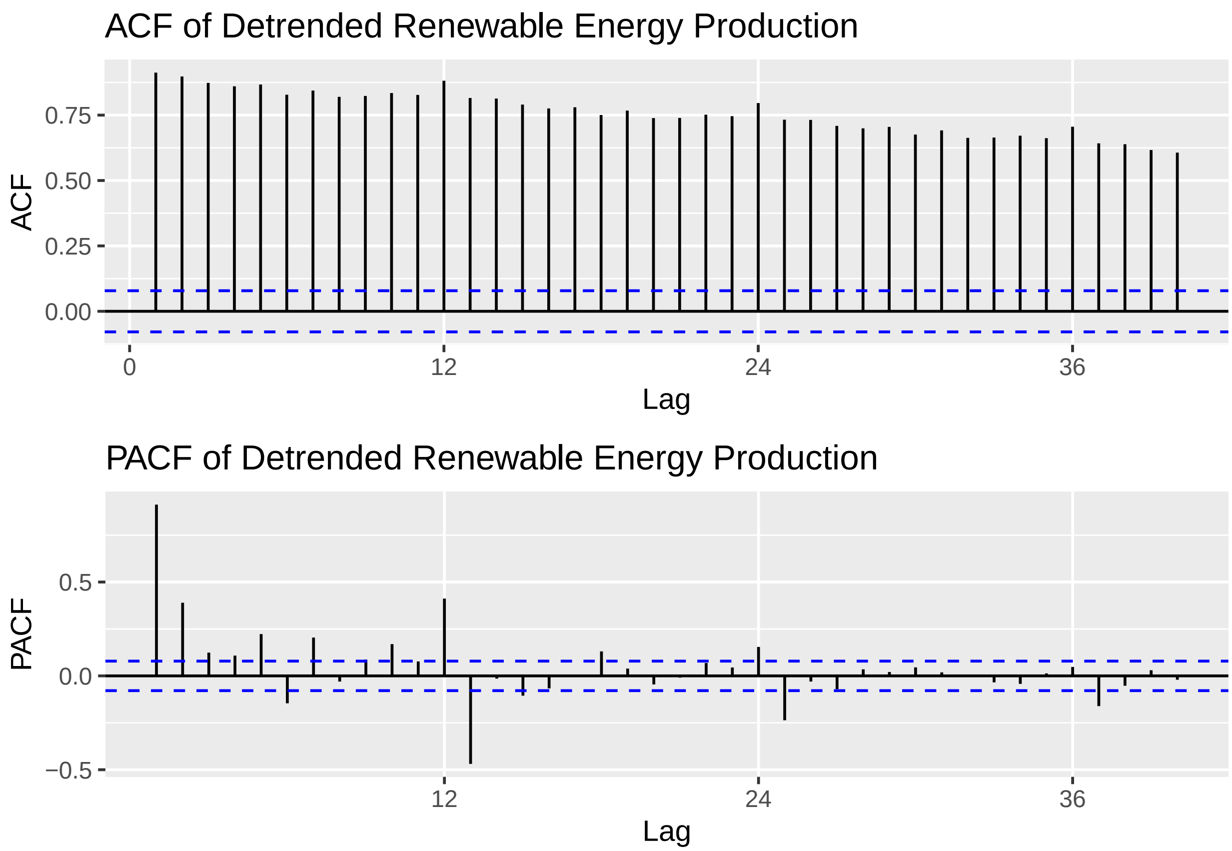
While **Detrended Hydroelectric Power Consumption** is similar to Q1, still fluctuating around a stable mean with seasonal patterns.

Q5

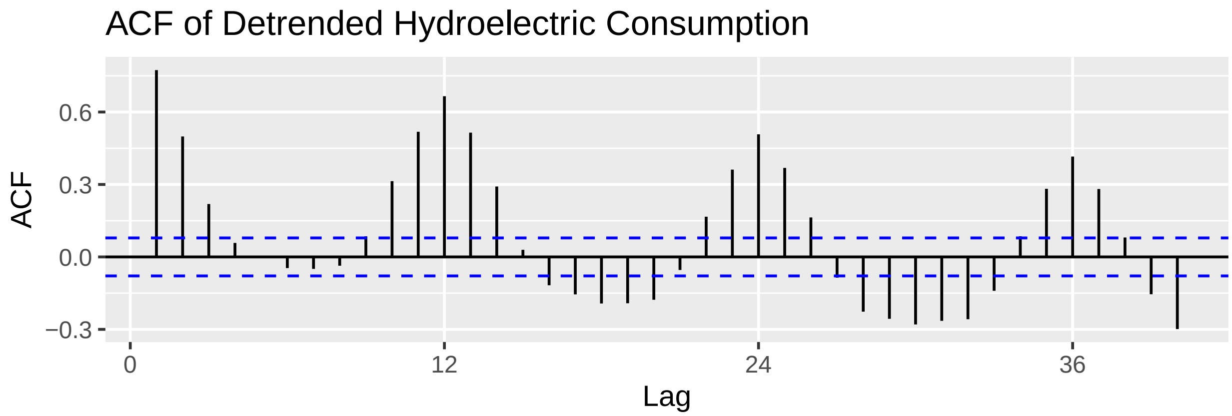
Plot ACF and PACF for the detrended series and compare with the plots from Q1. You may use plot_grid() again to get them side by side, but not mandatory. Did the plots change? How?

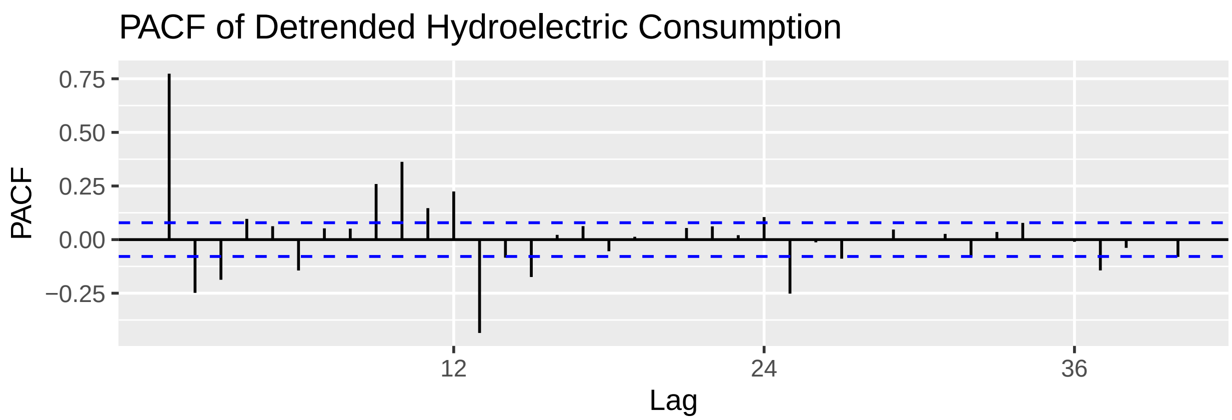
```r
#Function to create plots for the detrended series
plot_detrended_series <- function(ts_data, title) {
  acf_plot <- ggAcf(ts_data, lag.max=40) + ggtitle(paste("ACF of Detrended", title))
  pacf_plot <- ggPacf(ts_data, lag.max=40) + ggtitle(paste("PACF of Detrended", title))
```

```r
  plot_grid(acf_plot, pacf_plot, nrow=2)
}

#Plots for the detrended series
renewable_detrended_plot <- plot_detrended_series(ts_detrend_renewable, "Renewable Energy Production")
hydro_detrended_plot <- plot_detrended_series(ts_detrend_hydro, "Hydroelectric Consumption")

print(renewable_detrended_plot)
```




```r
print(hydro_detrended_plot)
```





For **Detrended Total Renewable Energy Production**, the ACF remains high, but compared to Q1, it shows a clearer seasonal pattern. Similarly, the number of significant lags in the PACF has increased, indicating stronger seasonal relationships.

For **Detrended Hydroelectric Power Consumption**, both the ACF and PACF are generally similar to Q1, but the peak values at certain lags are more obvious, showing a stronger seasonal pattern.

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 **Total Renewable Energy Production**, the time series plot shows some periodic fluctuations, but there is no clear seasonal pattern in the ACF plot, possibly a weak seasonal trend.

For **Hydroelectric Power Consumption**, the time series plot shows significant periodic fluctuations, and the ACF plot shows significant peaks at specific lags (like 12, 24, 36), indicating a strong seasonal trend.

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
#Create dummy variables
dummies <- seasonaldummy(ts_energy_data[,1])

#Fit a seasonal means model the two time series
seas_means_model_renewable <- lm(energy_data$`Total Renewable Energy Production` ~ dummies)
seas_means_model_hydro <- lm(energy_data$`Hydroelectric Power Consumption` ~ dummies)

#Print the summary of the regression
summary(seas_means_model_renewable)
```

```
##
## Call:
## lm(formula = energy_data$`Total Renewable Energy Production` ~
##     dummies)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -205.65  -91.59  -54.59  117.87  356.19
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  410.598     20.228  20.299   <2e-16 ***
## dummiesJan     1.973     28.468   0.069    0.945
## dummiesFeb   -34.348     28.468  -1.207    0.228
## dummiesMar     4.721     28.468   0.166    0.868
## dummiesApr    -7.896     28.468  -0.277    0.782
## dummiesMay     7.199     28.468   0.253    0.800
## dummiesJun    -3.394     28.468  -0.119    0.905
## dummiesJul     2.850     28.468   0.100    0.920
## dummiesAug    -4.430     28.468  -0.156    0.876
## dummiesSep   -29.037     28.468  -1.020    0.308
## dummiesOct   -20.205     28.606  -0.706    0.480
## dummiesNov   -20.706     28.606  -0.724    0.469
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 144.5 on 609 degrees of freedom
## Multiple R-squared:  0.008696,   Adjusted R-squared:  -0.009209
## F-statistic: 0.4857 on 11 and 609 DF,  p-value: 0.9126
```

```r
summary(seas_means_model_hydro)
```

```
##
## Call:
## lm(formula = energy_data$`Hydroelectric Power Consumption` ~
##     dummies)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -31.101  -6.241  -0.444   6.410  32.363
##
```

```
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   79.981      1.452  55.094  < 2e-16 ***
## dummiesJan     4.941      2.043   2.418 0.015880 *
## dummiesFeb    -2.513      2.043  -1.230 0.219219
## dummiesMar     7.053      2.043   3.452 0.000595 ***
## dummiesApr     5.399      2.043   2.642 0.008441 **
## dummiesMay    13.907      2.043   6.807 2.40e-11 ***
## dummiesJun    10.729      2.043   5.251 2.09e-07 ***
## dummiesJul     4.033      2.043   1.974 0.048845 *
## dummiesAug    -5.399      2.043  -2.643 0.008436 **
## dummiesSep   -16.596      2.043  -8.123 2.54e-15 ***
## dummiesOct   -16.370      2.053  -7.973 7.66e-15 ***
## dummiesNov   -10.805      2.053  -5.263 1.97e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.37 on 609 degrees of freedom
## Multiple R-squared:  0.4695, Adjusted R-squared:  0.4599
## F-statistic:    49 on 11 and 609 DF,  p-value: < 2.2e-16
```

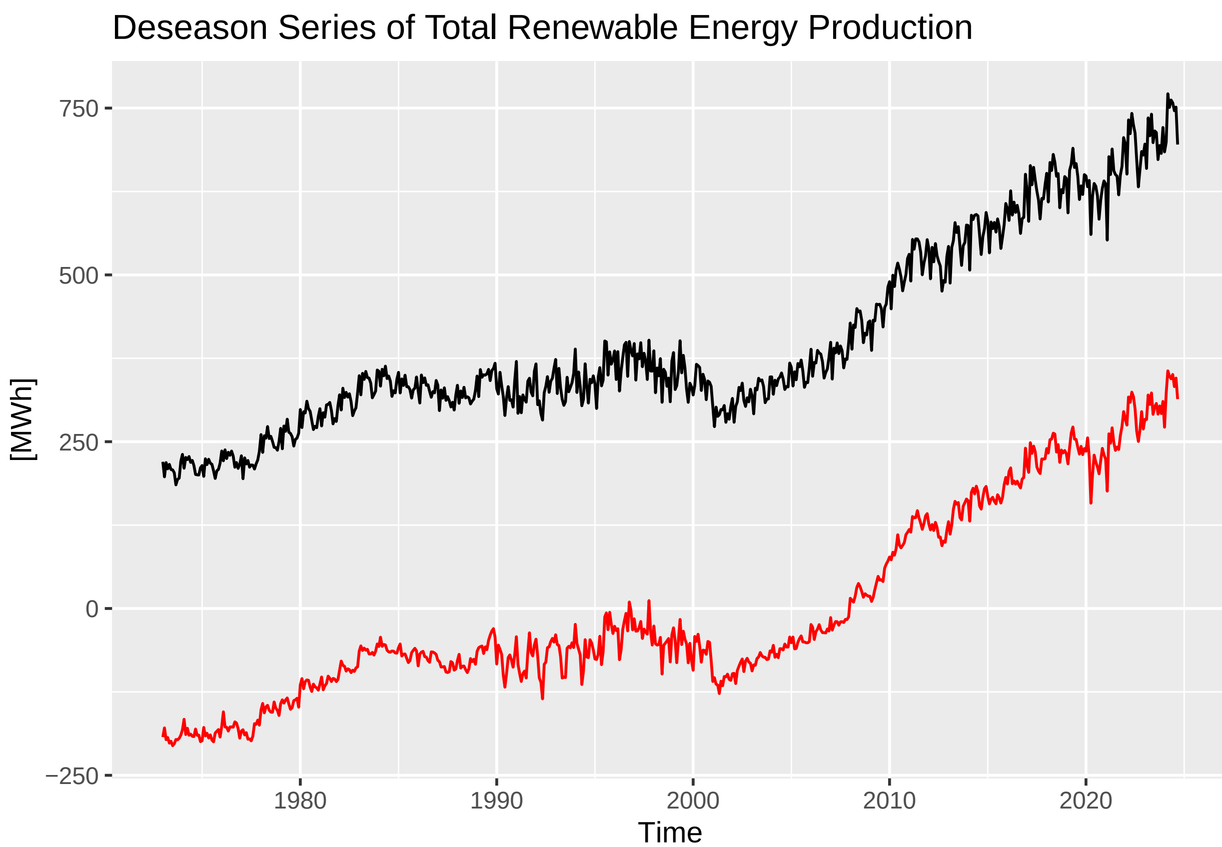
For the regression results of **Renewable Energy Production**, the p-values for all dummies are greater than 0.05 and the multiple R-squared (0.008696) is quite low, which means that seasonality is not significant. For the regression results of **Hydroelectric Power Consumption**, only the dummies of February have a p-value (0.219) greater than 0.05, while the dummies of the rest of the months are statistically significant, indicating a strong seasonal trend in this series.
This regression result is consistent with the observation of Q6.

Q8

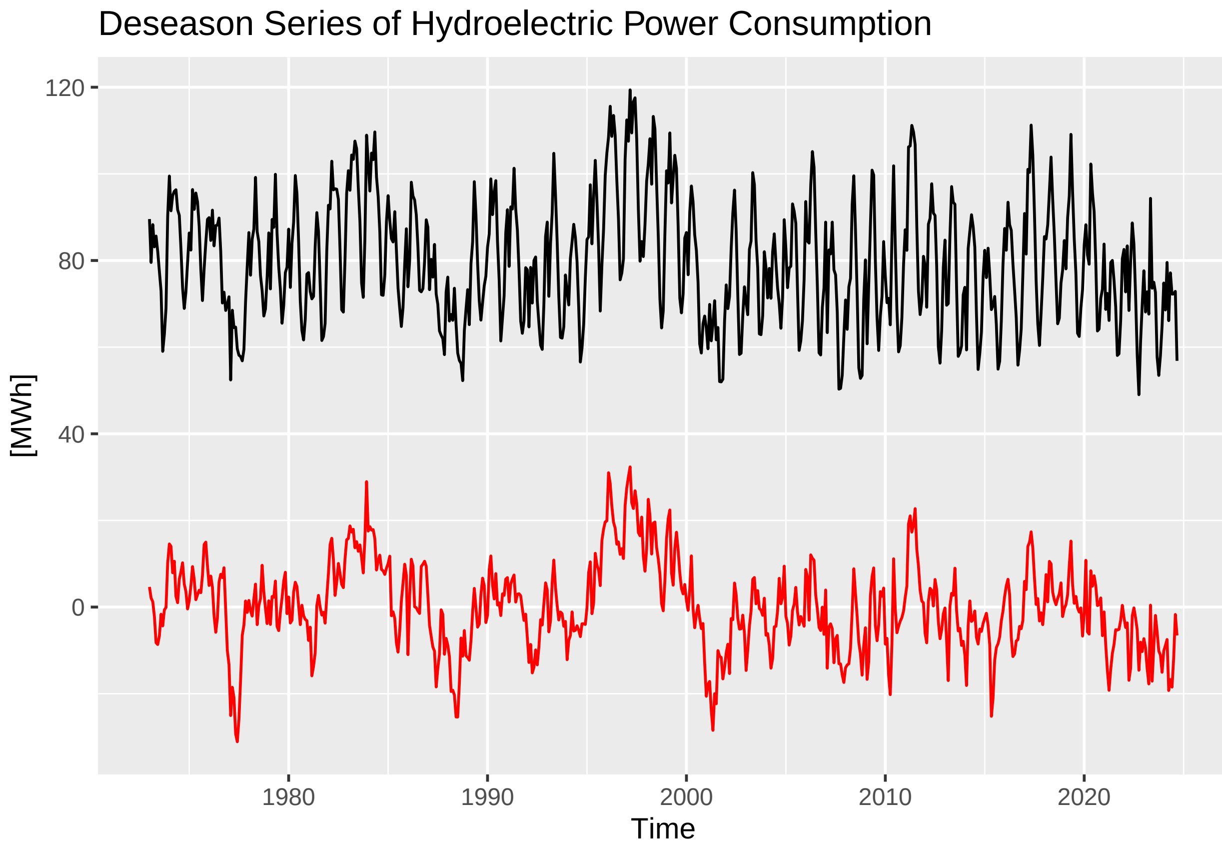
Use the regression coefficients from Q7 to deseason the series. Plot the deseason series and compare with the plots from part Q1. Did anything change?

```r
#Extract regression coefficients
beta_intercept_renewable <- seas_means_model_renewable$coefficients[1]
beta_coeff_renewable <- seas_means_model_renewable$coefficients[2:12]

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

#Deseason the series
renewable_seas_comp <- array(0,nobs)
hydro_seas_comp <- array(0,nobs)

for(i in 1:nobs) {
  renewable_seas_comp[i] <- beta_intercept_renewable + beta_coeff_renewable %*% dummies[i,]
  hydro_seas_comp[i] <- beta_intercept_hydro + beta_coeff_hydro %*% dummies[i,]
}

deseason_renewable <- ts_energy_data[,1] - renewable_seas_comp
deseason_hydro <- ts_energy_data[,2] - hydro_seas_comp

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

```r
ts_deseason_hydro <- ts(deseason_hydro, start=c(1973,1), frequency=12)

#Plot the deseason series
autoplot(ts_energy_data[,1]) +
  autolayer(ts_deseason_renewable, color="red") +
  ggtitle("Deseason Series of Total Renewable Energy Production") +
  ylab("[MWh]")
```




```r
autoplot(ts_energy_data[,2]) +
  autolayer(ts_deseason_hydro, color="red") +
  ggtitle("Deseason Series of Hydroelectric Power Consumption") +
  ylab("[MWh]")
```

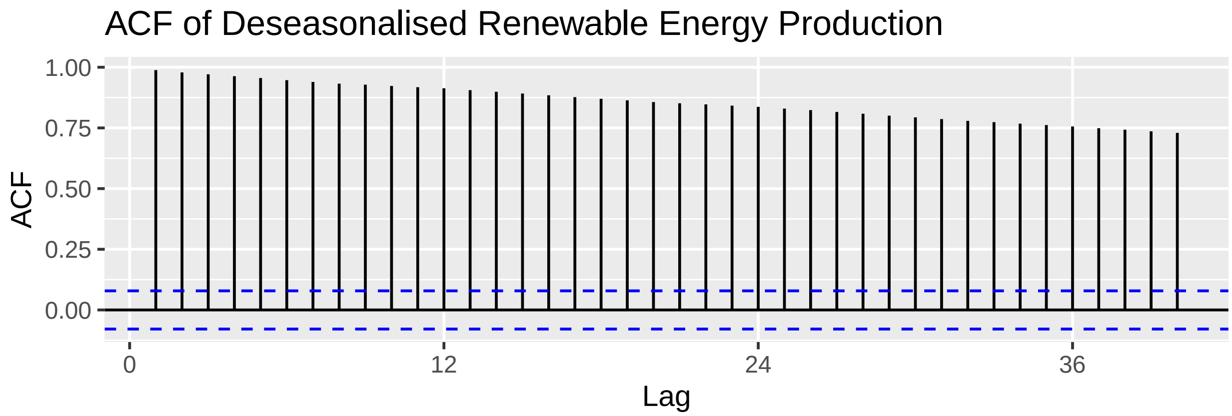


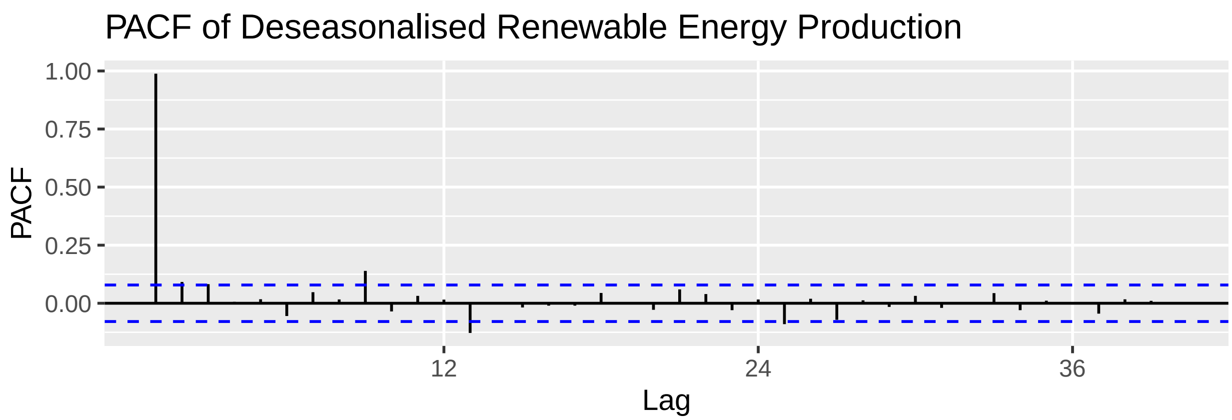
For **Deseasonalised Total Renewable Energy Production**, the long-term trend remains unchanged, with slightly reduced fluctuations, indicating little seasonal influence.
For **Deseasonalised Hydroelectric Power Consumption**, the overall trend is similar to Q1, but the fluctuation is significantly reduced, verifying a stronger seasonal effect.

Q9

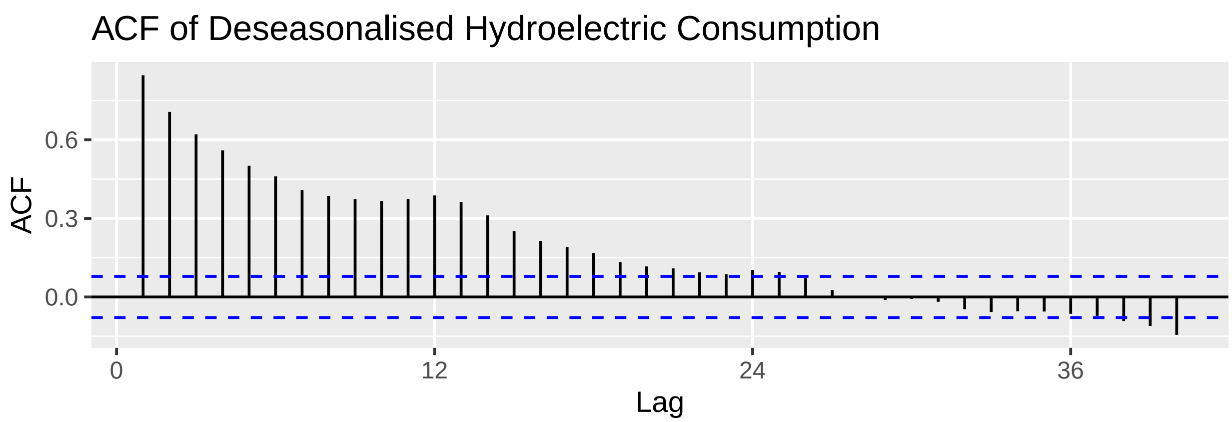
Plot ACF and PACF for the deseason series and compare with the plots from Q1. You may use plot_grid() again to get them side by side, but not mandatory. Did the plots change? How?

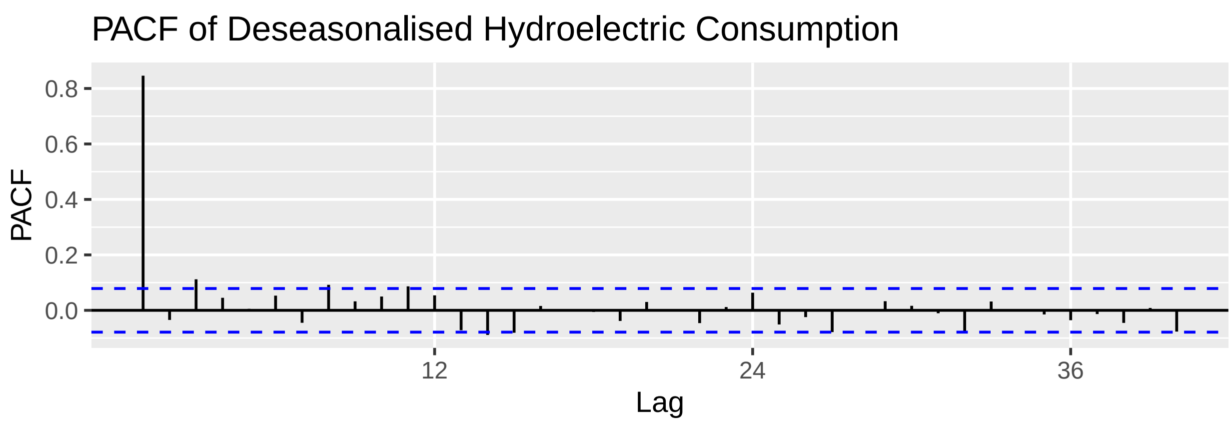
```r
#Function to create plots for the deseasonalised series
plot_deseason_series <- function(ts_data, title) {
  acf_plot <- ggAcf(ts_data, lag.max=40) + ggtitle(paste("ACF of Deseasonalised", title))
  pacf_plot <- ggPacf(ts_data, lag.max=40) + ggtitle(paste("PACF of Deseasonalised", title))

  plot_grid(acf_plot, pacf_plot, nrow=2)
}

#Plots for the deseasonalised series
renewable_deseason_plot <- plot_deseason_series(ts_deseason_renewable, "Renewable Energy Production")
hydro_deseason_plot <- plot_deseason_series(ts_deseason_hydro, "Hydroelectric Consumption")

print(renewable_deseason_plot)
```






```
print(hydro_deseason_plot)
```





For **Deseasonalised Total Renewable Energy Production**, the ACF is similar to that in Q1 and remains high, with even fewer significant lags in the PACF. The overall change is small, indicating that seasonal effects are weak and more influenced by long-term trends.
For **Deseasonalised Hydroelectric Power Consumption**, ACF shows an overall decreasing trend, and the original significant peaks of lag12, 24, and 36 in the ACF in Q1 are no longer significant. In PACF, nearly only lag 1 is significant, and significant lags are clearly reduced. This indicates that the seasonal effect of this series is strong.