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

Assignment 3 - Due date 02/03/26

Lingyue Hao

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_Consumpt The data comes from the US Energy Information and Administration and corresponds to the December 2025 Monthly Energy Review. This time 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.\

```
library(readxl)
library(openxlsx)
library(forecast)
```

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

```
library(tseries)
library(Kendall)
library(cowplot)
library(gridGraphics)
```

```
## Loading required package: grid
```

```r
library(ggplot2)
```

##Trend Component

Q1

For each series (Total Renewable Production and Hydroelectric Consumption) create three plots arranged in a row (side-by-side): (1) time series plot, (2) ACF, (3) PACF. Use cowplot::plot_grid() to place them in a grid.

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

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

```r
col_line <- read_excel(
  path = "../Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source.xlsx",
  sheet = "Monthly Data",
  skip = 10,
  n_max = 1,
  col_names = FALSE
)
```

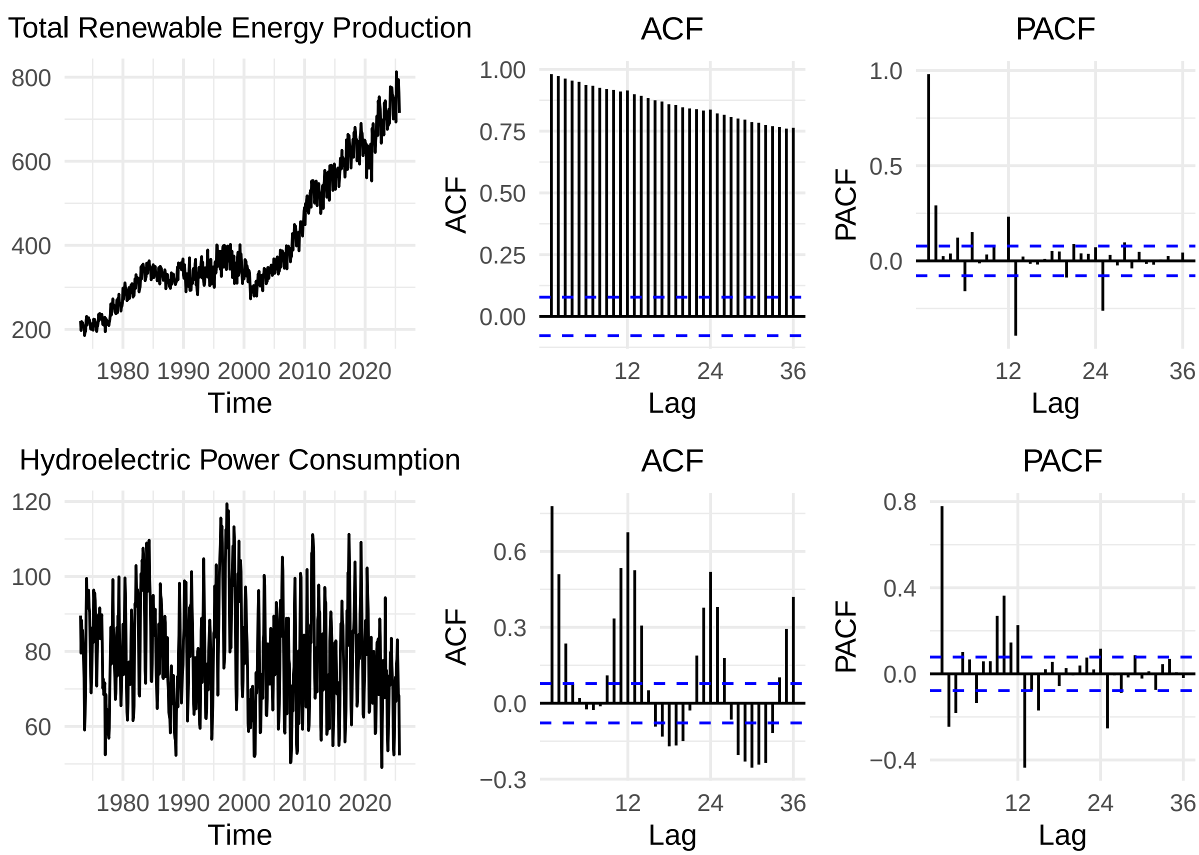
```
## New names:
## * `` -> `...1`
## * `` -> `...2`
## * `` -> `...3`
## * `` -> `...4`
## * `` -> `...5`
## * `` -> `...6`
## * `` -> `...7`
```

```
## * `` -> `...8`
## * `` -> `...9`
## * `` -> `...10`
## * `` -> `...11`
## * `` -> `...12`
## * `` -> `...13`
## * `` -> `...14`
```

```r
colnames(energy_data) <- as.character(col_line[1, ])

two_series_df <- energy_data[, c(
  "Total Renewable Energy Production",
  "Hydroelectric Power Consumption"
)]
two_series_df <- as.data.frame(two_series_df)

start_year <- 1973
start_month <- 1

trp_ts <- ts(as.numeric(two_series_df[["Total Renewable Energy Production"]]),
             start = c(start_year, start_month), frequency = 12)

hyd_ts <- ts(as.numeric(two_series_df[["Hydroelectric Power Consumption"]]),
             start = c(start_year, start_month), frequency = 12)
```

```r
plot_three_clean <- function(x_ts, series_name, lag_max = 36) {
  p_ts <- autoplot(x_ts) +
    labs(title = series_name,
         x = "Time", y = "") +
    theme_minimal(base_size = 11) +
    theme(
      plot.title = element_text(
        hjust = 0.5,
        size = 11,
        face = "plain"
      )
    )

  p_acf <- ggAcf(x_ts, lag.max = lag_max) +
    labs(title = "ACF") +
    theme_minimal(base_size = 11) +
    theme(
      plot.title = element_text(
        hjust = 0.5,
        size = 12,
        face = "plain"
      )
    )

  p_pacf <- ggPacf(x_ts, lag.max = lag_max) +
    labs(title = "PACF") +
    theme_minimal(base_size = 11) +
    theme(
      plot.title = element_text(
```

```
        hjust = 0.5,
        size = 12,
        face = "plain"
      )
    )

  cowplot::plot_grid(p_ts, p_acf, p_pacf, nrow = 1, rel_widths = c(1.2, 1, 1))
}

row_trp <- plot_three_clean(
  trp_ts,
  "Total Renewable Energy Production"
)

row_hyd <- plot_three_clean(
  hyd_ts,
  "Hydroelectric Power Consumption"
)
print(cowplot::plot_grid(row_trp, row_hyd, ncol = 1))
```



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 shows a clear upward trend over time, which indicates strong long-term growth. Because in the ACF plot, autocorrelations remain high and decay slowly across many lags. In contrast, Hydroelectric Power Consumption does not exhibit a clear long-term trend. The series fluctuates around a relatively stable level over time. Its ACF shows significant correlations at specific lags, which suggests seasonal or cyclical behavior rather than a sustained upward or downward trend.

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(two_series_df)

t <- 1:nobs

linear_trp <- lm(two_series_df$`Total Renewable Energy Production` ~ t)
summary(linear_trp)
```

```
##
## Call:
## lm(formula = two_series_df$`Total Renewable Energy Production` ~
##     t)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -154.81  -39.55   12.52   41.49  171.15
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 171.44868    5.11085   33.55   <2e-16 ***
## t             0.74999    0.01397   53.69   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 64.22 on 631 degrees of freedom
## Multiple R-squared:  0.8204, Adjusted R-squared:  0.8201
## F-statistic:  2883 on 1 and 631 DF,  p-value: < 2.2e-16
```

```r
trp_beta0 <- as.numeric(linear_trp$coefficients[1])
trp_beta1 <- as.numeric(linear_trp$coefficients[2])

linear_hyd <- lm(two_series_df$`Hydroelectric Power Consumption` ~ t)
summary(linear_hyd)
```

```
##
## Call:
## lm(formula = two_series_df$`Hydroelectric Power Consumption` ~
##     t)
##
## Residuals:
```

```
##      Min       1Q  Median       3Q      Max
## -30.190 -10.214  -0.715   8.909  39.723
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 83.223802   1.110552  74.939  < 2e-16 ***
## t           -0.012199   0.003035  -4.019 6.55e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.95 on 631 degrees of freedom
## Multiple R-squared:  0.02496,    Adjusted R-squared:  0.02342
## F-statistic: 16.15 on 1 and 631 DF,  p-value: 6.547e-05
```

```r
hyd_beta0 <- as.numeric(linear_hyd$coefficients[1])
hyd_beta1 <- as.numeric(linear_hyd$coefficients[2])

coef_list <- list(
  trp = c(beta0 = trp_beta0, beta1 = trp_beta1),
  hyd = c(beta0 = hyd_beta0, beta1 = hyd_beta1)
)
coef_list
```

```
## $trp
##      beta0      beta1
## 171.448682   0.749989
##
## $hyd
##       beta0       beta1
## 83.22380232 -0.01219868
```

For Total Renewable Energy Production, the intercept (171.45) shows the starting level of production at the beginning of the sample. The slope is about 0.75, which means that total renewable energy production increases by around 0.75 units each month on average. This indicates a clear upward trend over time. The p-value for the slope is smaller than 2e-16, so this increase is statistically significant and very unlikely to be due to random chance.

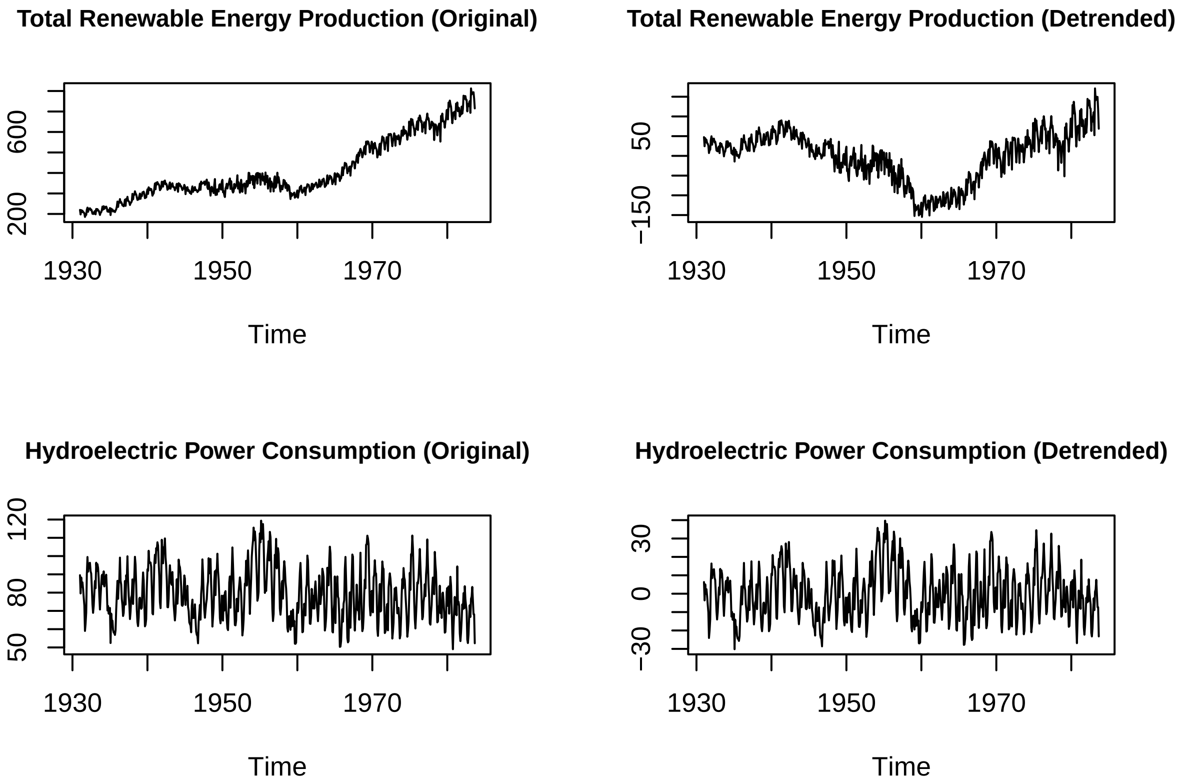
For Hydroelectric Power Consumption, the intercept (83.22) represents the initial level of consumption at the start of the period. The slope is approximately -0.012, meaning that hydroelectric power consumption decreases slightly by about 0.012 units per month. Although the change is small, it is consistent over time. The p-value for the slope is $6.547e-05$, which suggests that this downward trend is statistically significant, even though the size of the effect is relatively small.

Q4

Use the regression coefficients to detrend each series (subtract fitted linear trend). Plot detrended series and compare with the original time series from Q1. Describe what changed.

```r
fitted_trp <- coef_list$trp["beta0"] + coef_list$trp["beta1"] * t
fitted_hyd <- coef_list$hyd["beta0"] + coef_list$hyd["beta1"] * t

detrend_trp <- two_series_df$`Total Renewable Energy Production` - fitted_trp
```

```r
detrend_hyd <- two_series_df$`Hydroelectric Power Consumption` - fitted_hyd

ts_trp_det <- ts(detrend_trp, start = c(1931, 1), frequency = 12)
ts_hyd_det <- ts(detrend_hyd, start = c(1931, 1), frequency = 12)

par(mfrow = c(2, 2))

plot(ts(two_series_df$`Total Renewable Energy Production`,
        start = c(1931, 1), frequency = 12),
     main = "Total Renewable Energy Production (Original)",
     ylab = "", xlab = "Time",
     cex.main = 0.9)

plot(ts_trp_det,
     main = "Total Renewable Energy Production (Detrended)",
     ylab = "", xlab = "Time",
     cex.main = 0.9)

plot(ts(two_series_df$`Hydroelectric Power Consumption`,
        start = c(1931, 1), frequency = 12),
     main = "Hydroelectric Power Consumption (Original)",
     ylab = "", xlab = "Time",
     cex.main = 0.9)

plot(ts_hyd_det,
     main = "Hydroelectric Power Consumption (Detrended)",
     ylab = "", xlab = "Time",
     cex.main = 0.9)
```



After detrending, the long-term linear trend is removed from both series. For Total Renewable Energy Production, the strong upward growth visible in the original series disappears, and the detrended series fluctuates around zero. For Hydroelectric Power Consumption, the original series already shows no strong trend, so detrending mainly recenters the data around zero while preserving its variability and seasonal patterns.

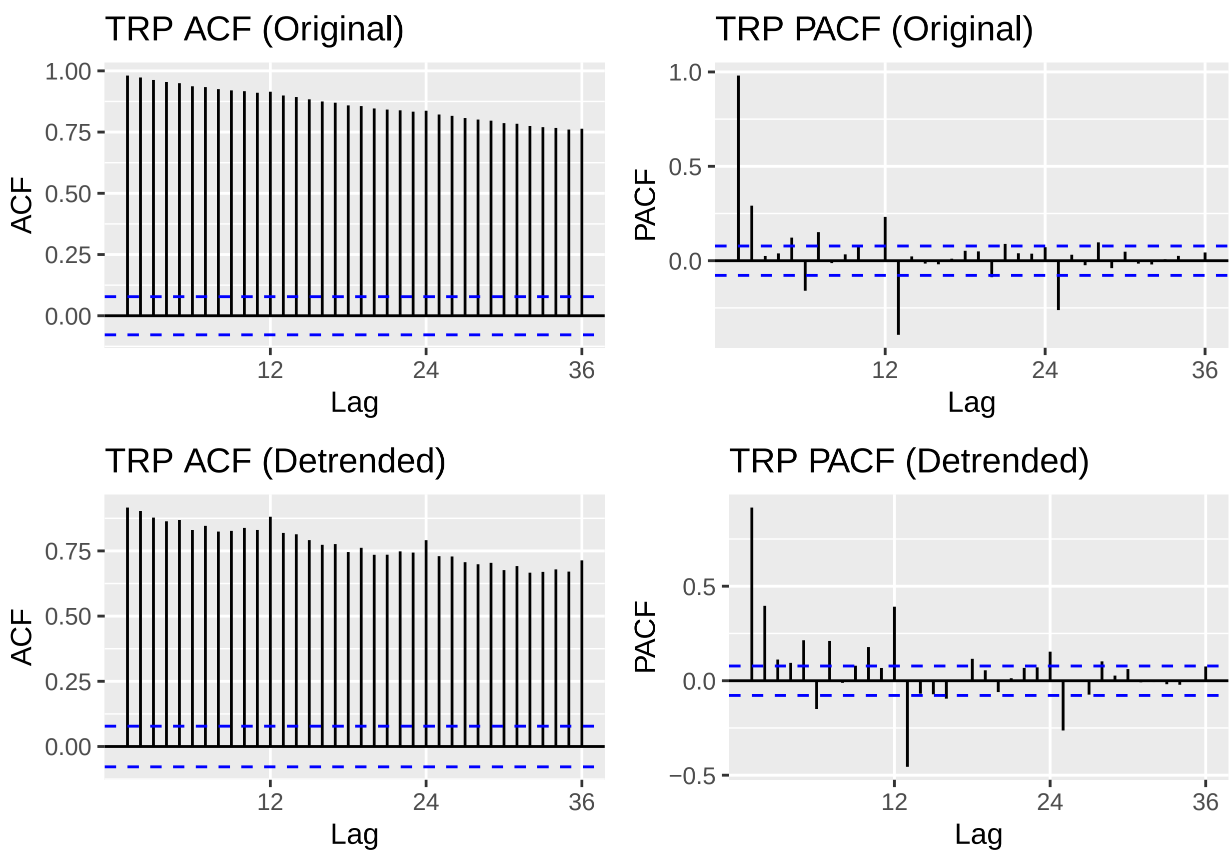
Q5

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 to make it easier to compare. Did the plots change? How?

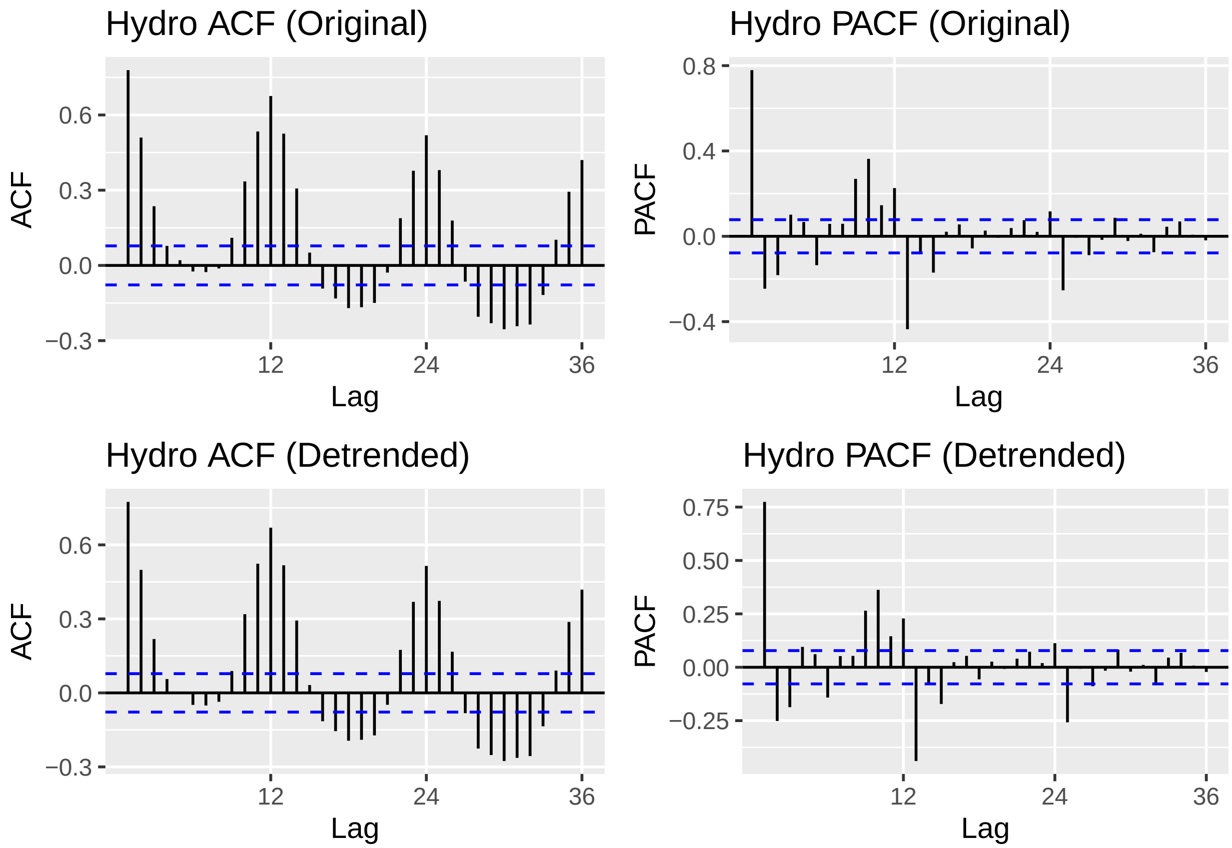
"TRP" in the plot below stands for Total Renewable Energy Production, "Hydro" in the plot below stands for Hydroelectric Power Consumption.

```
p1_o <- ggAcf(trp_ts, lag.max = 36) + ggtitle("TRP ACF (Original)")
p2_o <- ggPacf(trp_ts, lag.max = 36) + ggtitle("TRP PACF (Original)")
p3_o <- ggAcf(hyd_ts, lag.max = 36) + ggtitle("Hydro ACF (Original)")
p4_o <- ggPacf(hyd_ts, lag.max = 36) + ggtitle("Hydro PACF (Original)")

p1_d <- ggAcf(ts_trp_det, lag.max = 36) + ggtitle("TRP ACF (Detrended)")
p2_d <- ggPacf(ts_trp_det, lag.max = 36) + ggtitle("TRP PACF (Detrended)")
p3_d <- ggAcf(ts_hyd_det, lag.max = 36) + ggtitle("Hydro ACF (Detrended)")
p4_d <- ggPacf(ts_hyd_det, lag.max = 36) + ggtitle("Hydro PACF (Detrended)")

plot_grid(
  p1_o, p2_o,
```

```
  p1_d, p2_d,
  ncol = 2
)
```




```
plot_grid(
  p3_o, p4_o,
  p3_d, p4_d,
  ncol = 2
)
```



After detrending, the ACF and PACF plots do change compared to Q1, but the change is different for the two series. For Total Renewable Energy Production, removing the linear trend reduces part of the long-term increase, but the ACF still shows strong autocorrelation across many lags. The PACF now has only a few noticeable spikes, which suggests that the dependence structure becomes simpler once the trend is removed. For Hydroelectric Power Consumption, the ACF and PACF look quite similar to those in Q1, with clear spikes at seasonal lags such as 12. This suggests that hydro consumption is mainly driven by seasonality rather than a linear trend, so detrending does not change its correlation structure very much.

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.

> Answer: From the plots, Hydroelectric Power Consumption shows clear seasonality. The time series has regular repeating ups and downs, and the ACF shows significant spikes at lags 12, 24, and 36, which is consistent with an annual seasonal pattern. In contrast, Total Renewable Energy Production is mainly dominated by a strong upward trend. Its ACF stays high and decays slowly across many lags, so it is not a clear seasonal pattern.

Q7

Use function *lm()* to fit a seasonal means model (i.e. using the seasonal dummies) to 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?

```
trp_dummies <- seasonaldummy(trp_ts)
trp_season_model <- lm(trp_ts ~ trp_dummies)
summary(trp_season_model)
```

```
##
## Call:
## lm(formula = trp_ts ~ trp_dummies)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -213.33  -97.36  -59.88  121.55  389.62
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     417.265     21.096  19.779   <2e-16 ***
## trp_dummiesJan    2.090     29.693   0.070    0.944
## trp_dummiesFeb  -34.524     29.693  -1.163    0.245
## trp_dummiesMar    5.956     29.693   0.201    0.841
## trp_dummiesApr   -6.900     29.693  -0.232    0.816
## trp_dummiesMay    8.162     29.693   0.275    0.784
## trp_dummiesJun   -2.231     29.693  -0.075    0.940
## trp_dummiesJul    3.864     29.693   0.130    0.897
## trp_dummiesAug   -3.978     29.693  -0.134    0.893
## trp_dummiesSep  -29.033     29.693  -0.978    0.329
## trp_dummiesOct  -19.937     29.834  -0.668    0.504
## trp_dummiesNov  -20.617     29.834  -0.691    0.490
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 152.1 on 621 degrees of freedom
## Multiple R-squared:  0.008243,   Adjusted R-squared:  -0.009324
## F-statistic: 0.4692 on 11 and 621 DF,  p-value: 0.9223
```

```
hyd_dummies <- seasonaldummy(hyd_ts)
hyd_season_model <- lm(hyd_ts ~ hyd_dummies)
summary(hyd_season_model)
```

```
##
## Call:
## lm(formula = hyd_ts ~ hyd_dummies)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -30.895  -6.368  -0.595   6.213  32.557
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)      79.724      1.436  55.511  < 2e-16 ***
## hyd_dummiesJan    4.951      2.021   2.449 0.014591 *
## hyd_dummiesFeb   -2.415      2.021  -1.195 0.232608
## hyd_dummiesMar    7.116      2.021   3.520 0.000463 ***
## hyd_dummiesApr    5.614      2.021   2.777 0.005649 **
## hyd_dummiesMay   14.080      2.021   6.965 8.38e-12 ***
## hyd_dummiesJun   10.780      2.021   5.333 1.36e-07 ***
## hyd_dummiesJul    4.003      2.021   1.980 0.048091 *
## hyd_dummiesAug   -5.320      2.021  -2.632 0.008710 **
## hyd_dummiesSep  -16.598      2.021  -8.211 1.28e-15 ***
## hyd_dummiesOct  -16.329      2.031  -8.040 4.56e-15 ***
## hyd_dummiesNov  -10.782      2.031  -5.308 1.54e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.36 on 621 degrees of freedom
## Multiple R-squared:  0.4714, Adjusted R-squared:  0.4621
## F-statistic: 50.35 on 11 and 621 DF,  p-value: < 2.2e-16
```

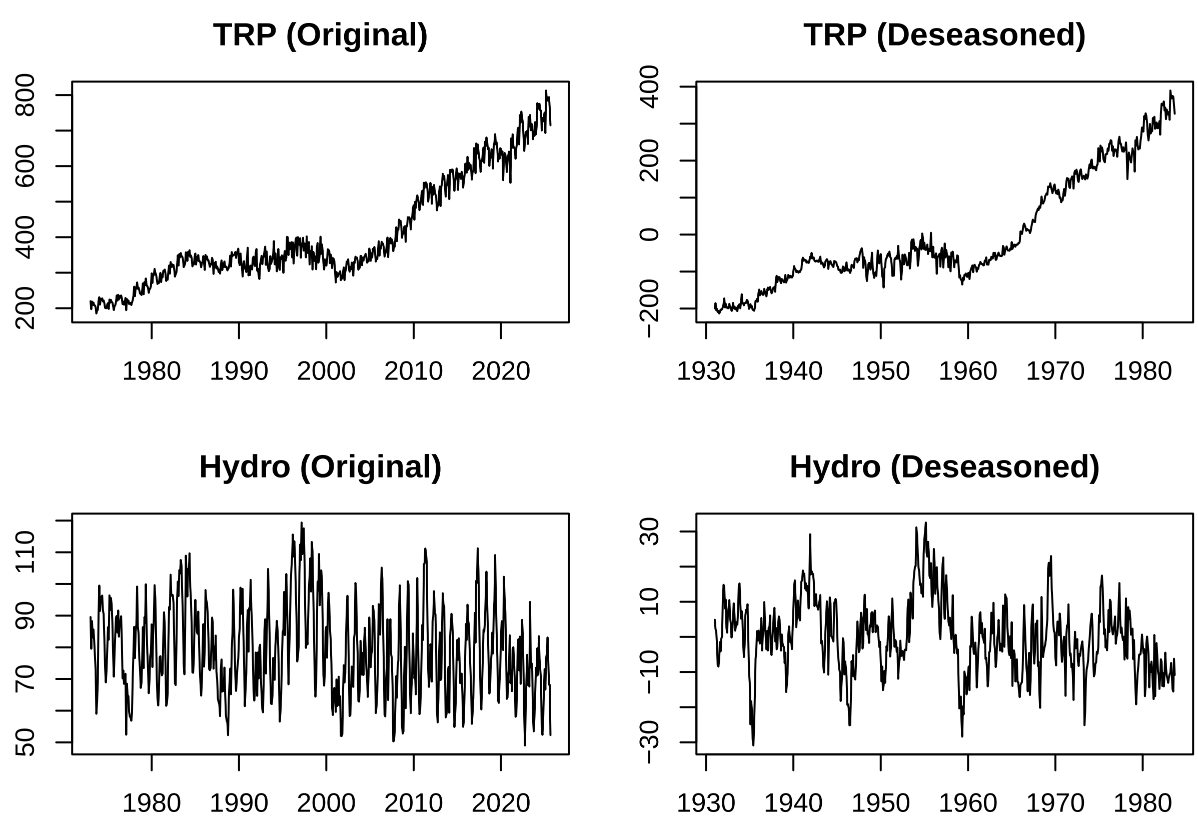
For Total Renewable Energy Production, none of the monthly dummy variables are statistically significant: all of their p-values are very large (above 0.05), and the overall F-test also has a high p-value (0.9223). This means we fail to reject the null hypothesis that all monthly effects are zero, so there is no evidence of a seasonal pattern in this series. In contrast, for Hydroelectric Power Consumption, many monthly dummy variables have very small p-values, and the overall F-test is highly significant. This provides strong statistical evidence of seasonality. So hydroelectric power consumption shows a clear seasonal trend, while total renewable energy production does not. These results are consistent with my answer to 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?

```
#TRP
trp_beta_int <- trp_season_model$coefficients[1]
trp_beta_coeff <- trp_season_model$coefficients[2:12]

nobs_trp <- length(trp_ts)
trp_season_component <- array(0, nobs_trp)
for(i in 1:nobs_trp){
  trp_season_component[i] <- trp_beta_int + trp_beta_coeff %*% trp_dummies[i, ]
}

trp_deseason <- trp_ts - trp_season_component

trp_deseason_ts <- ts(trp_deseason,
                      start = c(1931, 1),
                      frequency = 12)

#HYD
hyd_beta_int <- hyd_season_model$coefficients[1]
hyd_beta_coeff <- hyd_season_model$coefficients[2:12]
```

```
nobs_hyd <- length(hyd_ts)
hyd_season_component <- array(0, nobs_hyd)
for(i in 1:nobs_hyd){
  hyd_season_component[i] <- hyd_beta_int + hyd_beta_coeff %*% hyd_dummies[i, ]
}

hyd_deseason <- hyd_ts - hyd_season_component

hyd_deseason_ts <- ts(hyd_deseason,
                      start = c(1931, 1),
                      frequency = 12)


par(mfrow = c(2, 2), mar = c(3, 3, 3, 1))

plot(trp_ts, main = "TRP (Original)", xlab = "Time", ylab = "")
plot(trp_deseason_ts, main = "TRP (Deseasoned)", xlab = "Time", ylab = "")

plot(hyd_ts, main = "Hydro (Original)", xlab = "Time", ylab = "")
plot(hyd_deseason_ts, main = "Hydro (Deseasoned)", xlab = "Time", ylab = "")
```



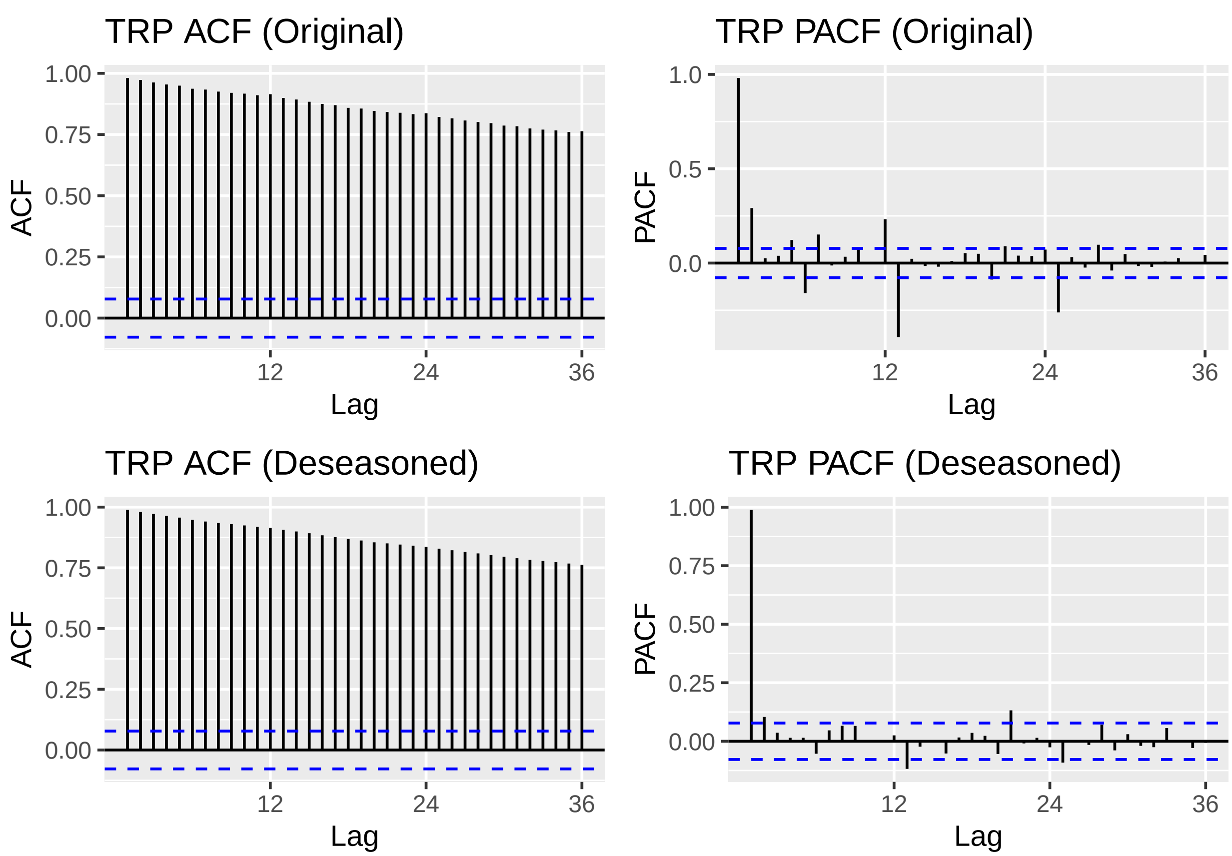
Yes, but the change is very different for the two series.For Total Renewable Energy Production, the deseasoned series looks almost the same as the original. The overall pattern and long-term movement remain unchanged. For Hydroelectric Power Consumption, the change is much more noticeable. After deseasoning, the regular monthly up-and-down pattern is largely removed, and the series fluctuates more evenly around zero.

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. Did the plots change? How?

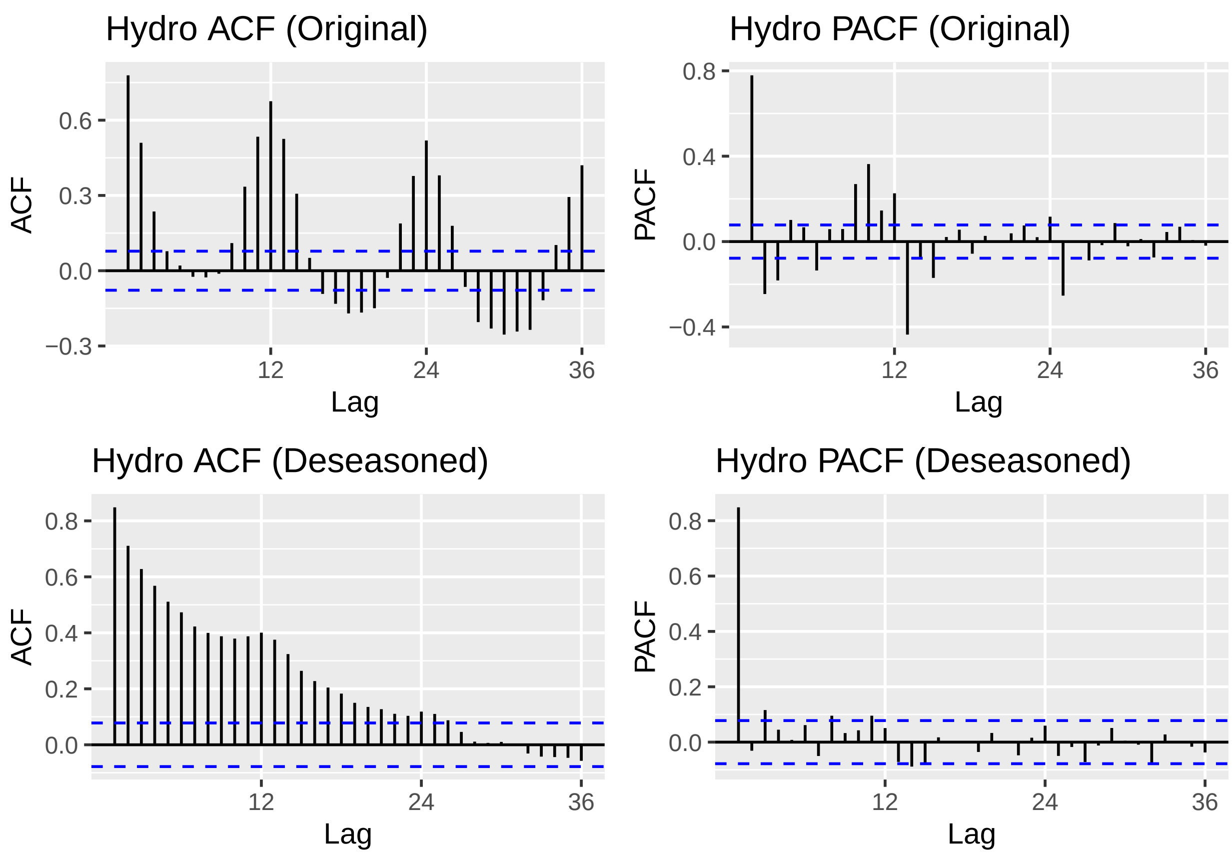
“TRP” in the plot below stands for Total Renewable Energy Production, “Hydro” in the plot below stands for Hydroelectric Power Consumption.

```
# TRP
p1 <- ggAcf(trp_ts, lag.max = 36) +
  ggtitle("TRP ACF (Original)")

p2 <- ggPacf(trp_ts, lag.max = 36) +
  ggtitle("TRP PACF (Original)")

p3 <- ggAcf(trp_deseason_ts, lag.max = 36) +
  ggtitle("TRP ACF (Deseasoned)")

p4 <- ggPacf(trp_deseason_ts, lag.max = 36) +
  ggtitle("TRP PACF (Deseasoned)")

plot_grid(p1, p2, p3, p4, ncol = 2)
```




```
# Hydro
p5 <- ggAcf(hyd_ts, lag.max = 36) +
  ggtitle("Hydro ACF (Original)")
```

```r
p6 <- ggPacf(hyd_ts, lag.max = 36) +
  ggtitle("Hydro PACF (Original)")

p7 <- ggAcf(hyd_deseason_ts, lag.max = 36) +
  ggtitle("Hydro ACF (Deseasoned)")

p8 <- ggPacf(hyd_deseason_ts, lag.max = 36) +
  ggtitle("Hydro PACF (Deseasoned)")

plot_grid(p5, p6, p7, p8, ncol = 2)
```



Yes, the plots do change after deseasoning, especially for the Hydroelectric Power Consumption series. After removing the seasonal component, the Hydroelectric Power Consumption no longer shows strong spikes at seasonal lags (like 12, 24), and the correlations die out much faster. The hydro PACF also becomes cleaner, with fewer significant spikes, which means that most of the seasonal pattern has been removed. For Total Renewable Energy Production, the ACF still shows very strong persistence across many lags and the PACF looks similar to before, so deseasoning does not change it much.