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

Assignment 3 - Due date 02/03/26

Matthew Fischer

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_Consumption 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.\

```
#Load/install required package here
library(forecast)
```

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

```
library(tseries)
library(Kendall)
library("readxl")
library(ggplot2)
library(cowplot)

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

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

```
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`
```

```
colnames(energy_data) <- read_col_names

energy_subset <- data.frame(
  Total_Renewable_Energy_Production =
    energy_data$"Total Renewable Energy Production",
  Hydroelectric_Power_Consumption =
    energy_data$"Hydroelectric Power Consumption"
)

head(energy_subset)
```

```
##   Total_Renewable_Energy_Production Hydroelectric_Power_Consumption
## 1                           219.839                          89.562
## 2                           197.330                          79.544
## 3                           218.686                          88.284
## 4                           209.330                          83.152
## 5                           215.982                          85.643
## 6                           208.249                          82.060
```

```
energy_ts <- ts(
  energy_subset,
  start = c(1973, 1),
```

```r
  frequency = 12
)
```

##Trend Component

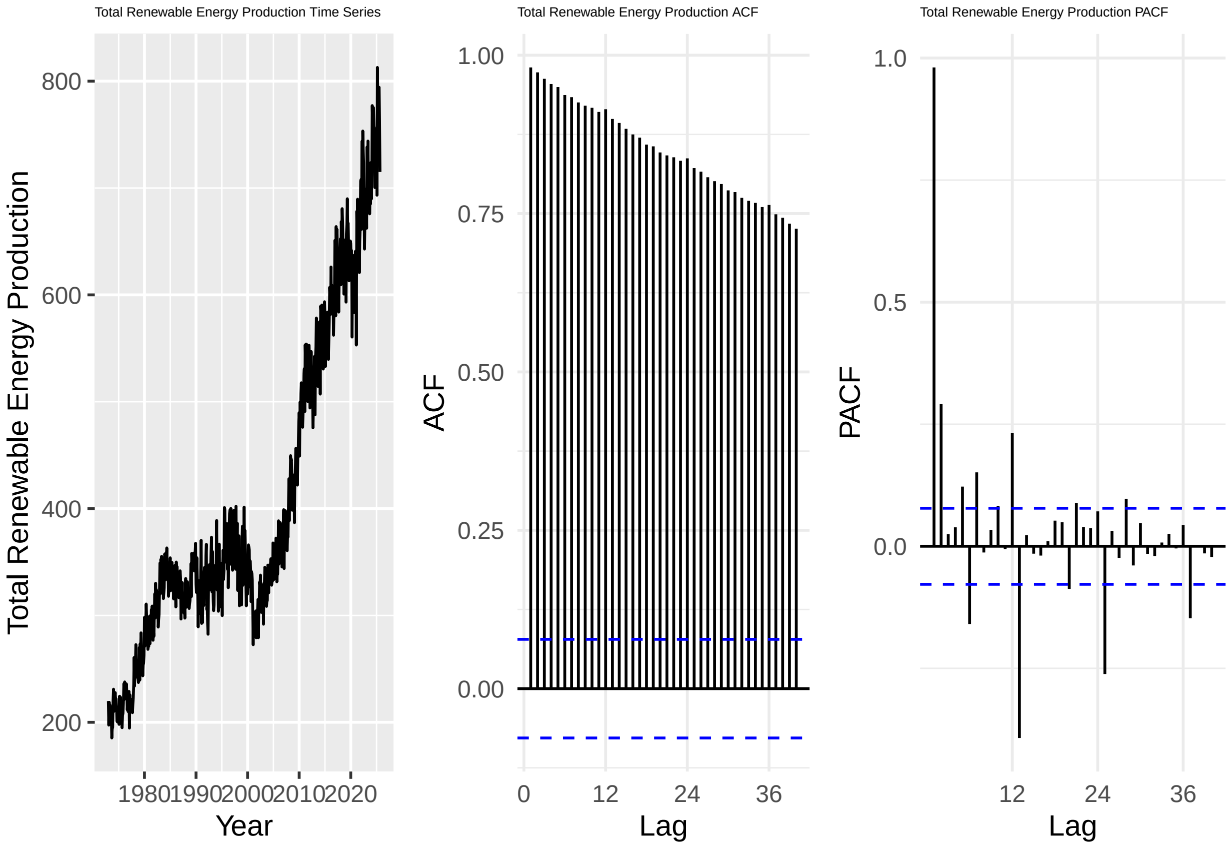
Q1

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
trp_ts <- autoplot(energy_ts[, 1], lag.max = 40, plot = TRUE) +
  ggtitle(paste("Total Renewable Energy Production Time Series")) +
  xlab("Year") + ylab("Total Renewable Energy Production") +
  theme(plot.title = element_text(size = 5))
```

```
## Warning in ggplot2::geom_line(na.rm = TRUE, ...): Ignoring unknown parameters:
## `lag.max` and `plot`
```

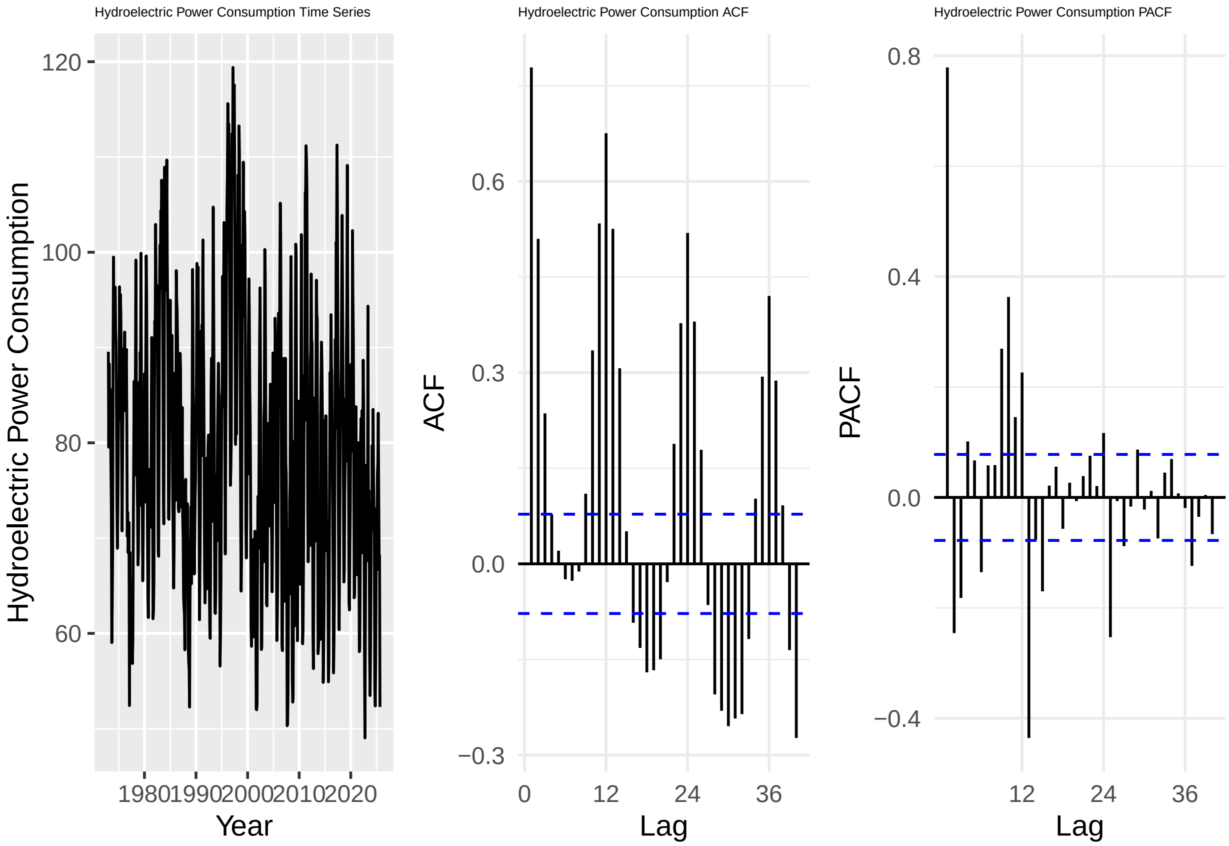
```r
trp_acf <- ggAcf(energy_ts[, 1], lag.max = 40, plot = TRUE) +
  ggtitle(paste("Total Renewable Energy Production ACF")) +
  theme_minimal() +
  theme(plot.title = element_text(size = 5))

trp_pacf <- ggPacf(energy_ts[, 1], lag.max = 40, plot = TRUE) +
  ggtitle(paste("Total Renewable Energy Production PACF")) +
  theme_minimal() +
  theme(plot.title = element_text(size = 5))

plot_grid(trp_ts, trp_acf, trp_pacf, nrow = 1)
```




```
hc_ts <- autoplot(energy_ts[, 2], lag.max = 40, plot = TRUE) +
  ggtitle(paste("Hydroelectric Power Consumption Time Series")) +
  xlab("Year") + ylab("Hydroelectric Power Consumption") +
  theme(plot.title = element_text(size = 5))
```

```
## Warning in ggplot2::geom_line(na.rm = TRUE, ...): Ignoring unknown parameters:
## `lag.max` and `plot`
```

```
hc_acf <- ggAcf(energy_ts[, 2], lag.max = 40, plot = TRUE) +
  ggtitle(paste("Hydroelectric Power Consumption ACF")) +
  theme_minimal() +
  theme(plot.title = element_text(size = 5))

hc_pacf <- ggPacf(energy_ts[, 2], lag.max = 40, plot = TRUE) +
  ggtitle(paste("Hydroelectric Power Consumption PACF")) +
  theme_minimal() +
  theme(plot.title = element_text(size = 5))

plot_grid(hc_ts, hc_acf, hc_pacf, nrow = 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?

> Answer: Total Renewable Energy Production has a definite positive trend while Hydroelectric Power Consumption does not appear to have a 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
t <- time(energy_ts)

trp_lm <- lm(energy_ts[, 1] ~ t)
hc_lm <- lm(energy_ts[, 2] ~ t)

summary(trp_lm)
```

```
##
## Call:
## lm(formula = energy_ts[, 1] ~ 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) -1.758e+04  3.351e+02  -52.47   <2e-16 ***
## t            9.000e+00  1.676e-01   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
summary(hc_lm)
```

```
##
## Call:
## lm(formula = energy_ts[, 2] ~ 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) 372.02750   72.82164   5.109 4.30e-07 ***
## t            -0.14638    0.03642  -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
```

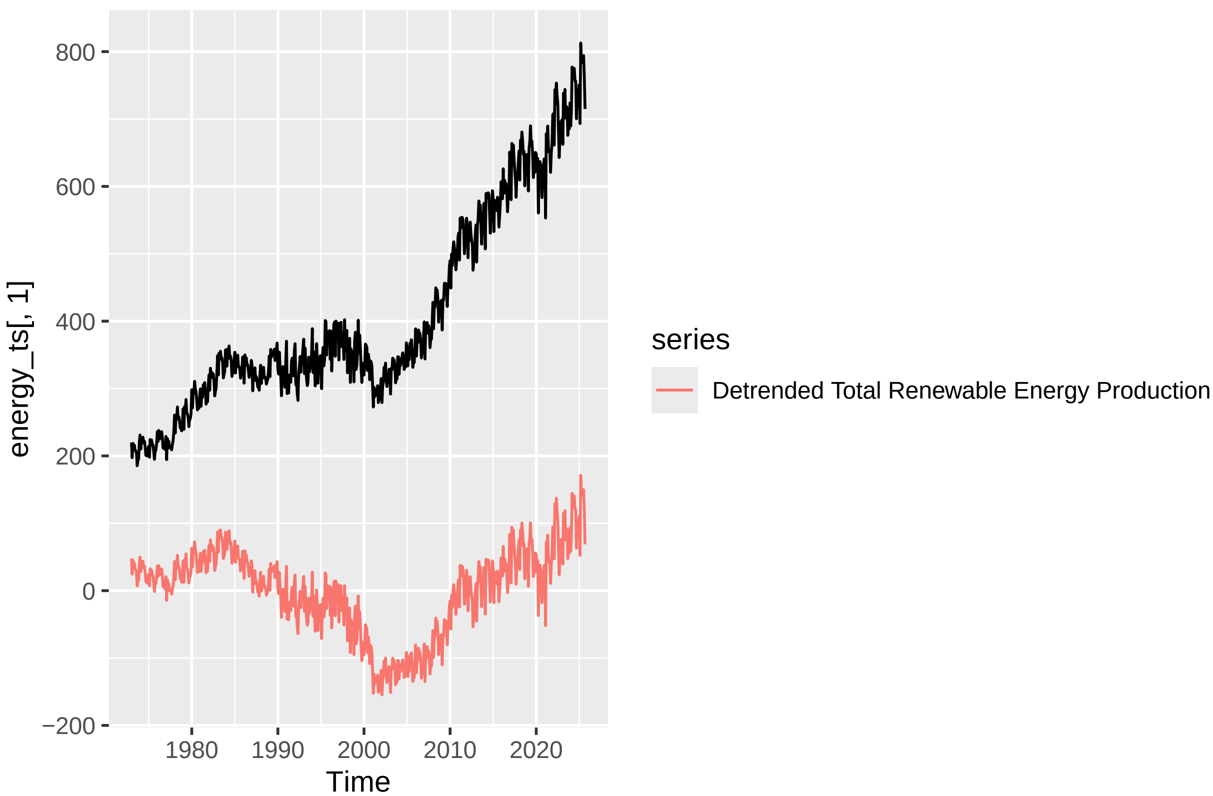
```r
trp_int <- as.numeric(trp_lm$coefficients[1])
trp_slope <- as.numeric(trp_lm$coefficients[2])

hc_int <- as.numeric(hc_lm$coefficients[1])
hc_slope <- as.numeric(hc_lm$coefficients[2])
```

> Answer: The Total Renewable Energy Production on average increases by 9 trillion Btu every year with a hypothetical value of -1758 Btu at time = 0. The Hydroelectric Power Consumption on average decreases by 0.146 trillion Btu every year with a hypothetical value of 372 Btu at time = 0.

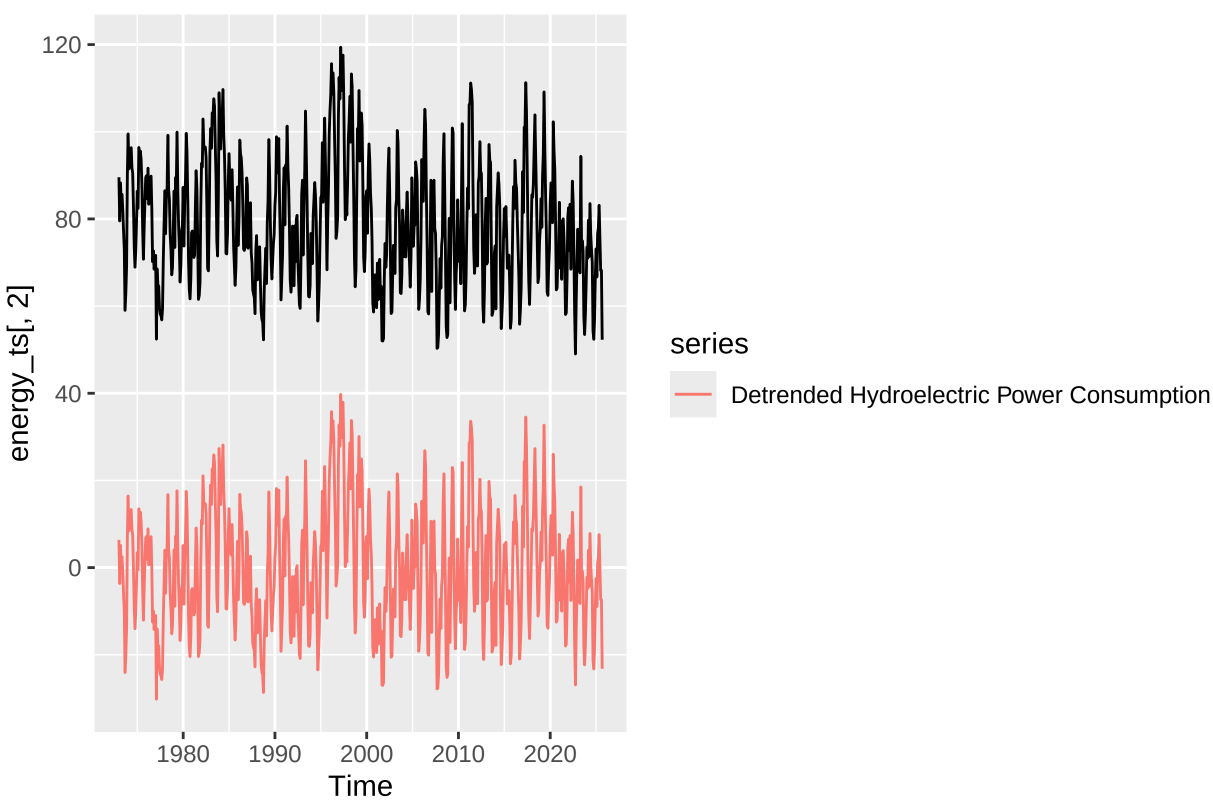
Q4

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
trp_detrended <- energy_ts[, 1] - (trp_int + (trp_slope * t))
hc_detrended <- energy_ts[, 2] - (hc_int + (hc_slope * t))

autoplot(energy_ts[, 1]) +
  autolayer(trp_detrended, series = "Detrended Total Renewable Energy Production")
```




```r
autoplot(energy_ts[, 2]) +
  autolayer(hc_detrended, series = "Detrended Hydroelectric Power Consumption")
```

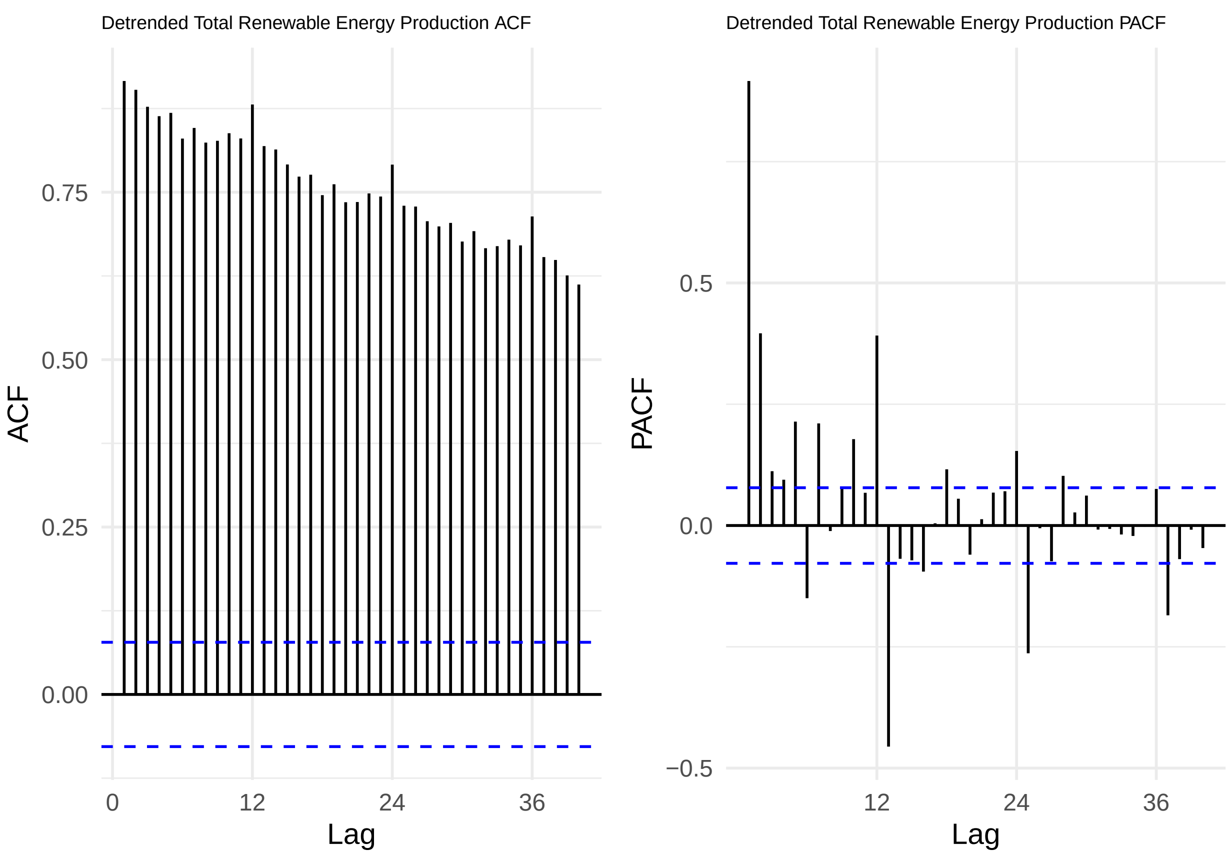


Answer: The detrended Total Renewable Energy Production is 150-650 trillion Btu lower than the original Total Renewable Energy Production and the trend is overall negative until 2000, at which point the trend becomes overall positive. The detrended Hydroelectric Power Consumption is about 80 trillion Btu lower than the original Hydroelectric Power Consumption, but other than that the graphs look very similar.

Q5

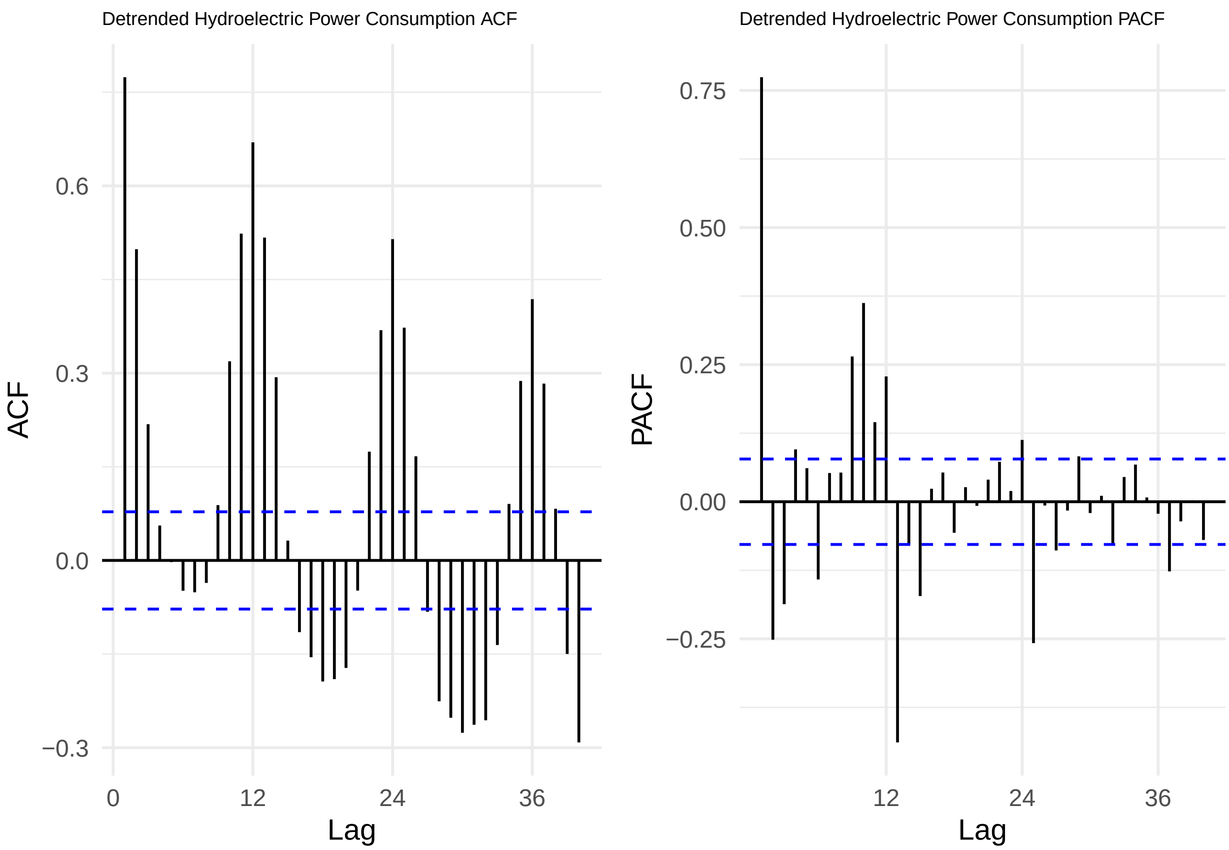
Plot ACF and PACF for the detrended series and compare with the plots from Q1. You may use plot_grid() again to get them side by side to make it easier to compare. Did the plots change? How?

```
trp_detrended_acf <- ggAcf(trp_detrended, lag.max = 40, plot = TRUE) +
  ggtitle(paste("Detrended Total Renewable Energy Production ACF")) +
  theme_minimal() +
  theme(plot.title = element_text(size = 7))

trp_detrended_pacf <- ggPacf(trp_detrended, lag.max = 40, plot = TRUE) +
  ggtitle(paste("Detrended Total Renewable Energy Production PACF")) +
  theme_minimal() +
  theme(plot.title = element_text(size = 7))

plot_grid(trp_detrended_acf, trp_detrended_pacf, nrow = 1)
```




```
hc_detrended_acf <- ggAcf(hc_detrended, lag.max = 40, plot = TRUE) +
  ggtitle(paste("Detrended Hydroelectric Power Consumption ACF")) +
  theme_minimal() +
  theme(plot.title = element_text(size = 7))

hc_detrended_pacf <- ggPacf(hc_detrended, lag.max = 40, plot = TRUE) +
```

```
  ggtitle(paste("Detrended Hydroelectric Power Consumption PACF")) +
  theme_minimal() +
  theme(plot.title = element_text(size = 7))

plot_grid(hc_detrended_acf, hc_detrended_pacf, nrow = 1)
```



Answer: The detrended ACF for Total Renewable Energy Production is about 0.1 to 0.15 lower than the original ACF for Total Renewable Energy Production, but other than that they look similar. The detrended PACF for Total Renewable Energy Production and the original ACF and PACF for Hydroelectric Power Consumption look the same as their seasoned counterparts.

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: Only from looking at the time series and ACF plots, Total Renewable Energy Production does not seem to have much of a seasonal trend and Hydroelectric Power Consumption does seem to have a seasonal trend.

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?

```r
trp_detrended_ts <- ts(trp_detrended, frequency = 12, start = c(1931, 1))
hc_detrended_ts <- ts(hc_detrended, frequency = 12, start = c(1931, 1))

trp_dummies <- seasonaldummy(trp_detrended_ts)
hc_dummies <- seasonaldummy(hc_detrended_ts)

trp_seas_means_model <- lm(trp_detrended_ts~trp_dummies)
hc_seas_means_model <- lm(hc_detrended_ts~hc_dummies)

summary(trp_seas_means_model)
```

```
##
## Call:
## lm(formula = trp_detrended_ts ~ trp_dummies)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -153.09  -36.94   15.01   42.21  155.62
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)       7.320      8.763   0.835   0.4039
## trp_dummiesJan    5.840     12.334   0.473   0.6360
## trp_dummiesFeb  -31.525     12.334  -2.556   0.0108 *
## trp_dummiesMar    8.205     12.334   0.665   0.5061
## trp_dummiesApr   -5.400     12.334  -0.438   0.6617
## trp_dummiesMay    8.912     12.334   0.723   0.4703
## trp_dummiesJun   -2.231     12.334  -0.181   0.8565
## trp_dummiesJul    3.114     12.334   0.252   0.8008
## trp_dummiesAug   -5.478     12.334  -0.444   0.6571
## trp_dummiesSep  -31.283     12.334  -2.536   0.0114 *
## trp_dummiesOct  -18.437     12.393  -1.488   0.1373
## trp_dummiesNov  -19.867     12.393  -1.603   0.1094
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 63.19 on 621 degrees of freedom
## Multiple R-squared:  0.04701,    Adjusted R-squared:  0.03013
## F-statistic: 2.785 on 11 and 621 DF,  p-value: 0.00152
```

```r
summary(hc_seas_means_model)
```

```
##
## Call:
## lm(formula = hc_detrended_ts ~ hc_dummies)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -34.116  -5.871  -0.555   5.823  32.264
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)     0.3796     1.4030   0.271 0.786811
```

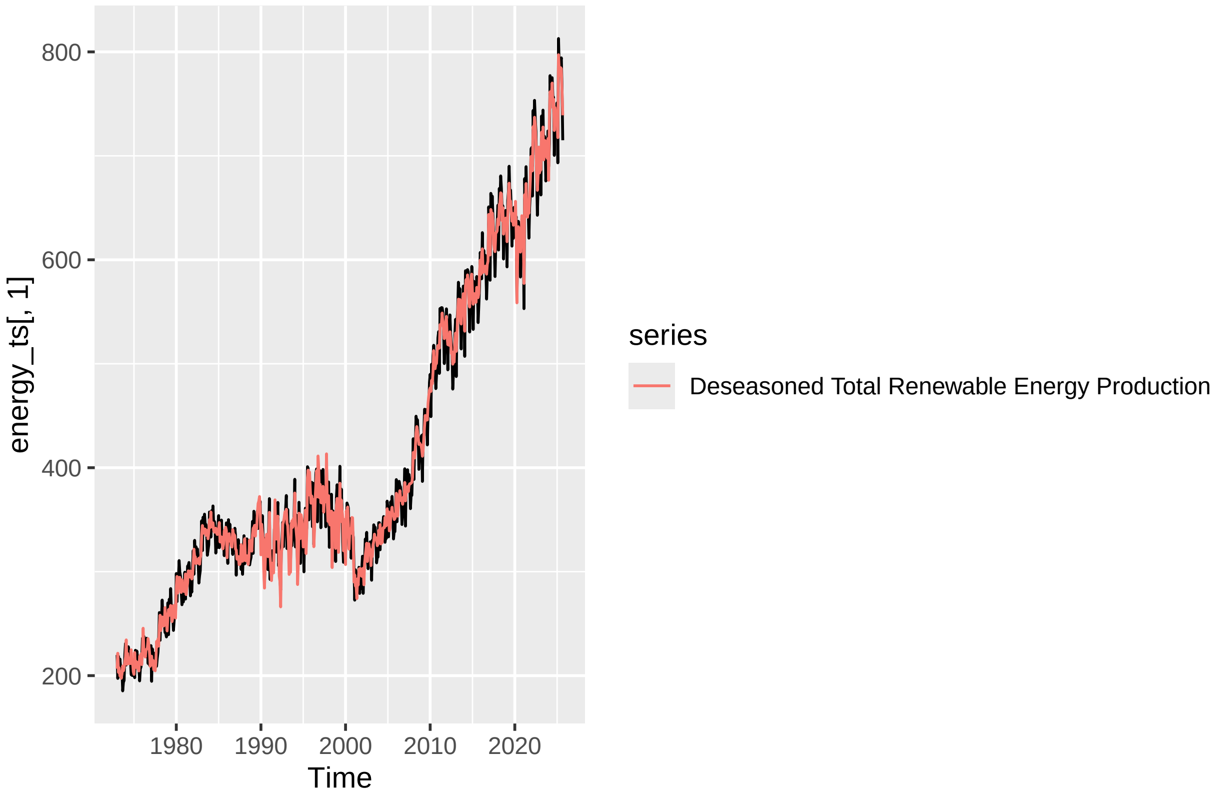
```
## hc_dummiesJan   4.8900     1.9747   2.476 0.013541 *
## hc_dummiesFeb  -2.4641     1.9747  -1.248 0.212573
## hc_dummiesMar   7.0794     1.9747   3.585 0.000364 ***
## hc_dummiesApr   5.5895     1.9747   2.830 0.004798 **
## hc_dummiesMay  14.0676     1.9747   7.124 2.92e-12 ***
## hc_dummiesJun  10.7799     1.9747   5.459 6.93e-08 ***
## hc_dummiesJul   4.0156     1.9747   2.033 0.042427 *
## hc_dummiesAug  -5.2952     1.9747  -2.681 0.007525 **
## hc_dummiesSep -16.5612     1.9747  -8.386 3.37e-16 ***
## hc_dummiesOct -16.3534     1.9841  -8.242 1.01e-15 ***
## hc_dummiesNov -10.7940     1.9841  -5.440 7.67e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.12 on 621 degrees of freedom
## Multiple R-squared:  0.4827, Adjusted R-squared:  0.4735
## F-statistic: 52.67 on 11 and 621 DF,  p-value: < 2.2e-16
```

> Answer: For the Total Renewable Energy Production, only the February and September dummies are statistically significant and the adjusted R-squared value is 0.030, meaning that only 3% of the variation is explained by seasonality. Therefore, there is no meaningful seasonal trend. For the Hydroelectric Power Consumption, most months are highly statistically significant and the adjusted R-squared value is 0.474, meaning that 47.4% of the variation can be explained by seasonality. Therefore, there is strong seasonality. These results match my answer to Q6.

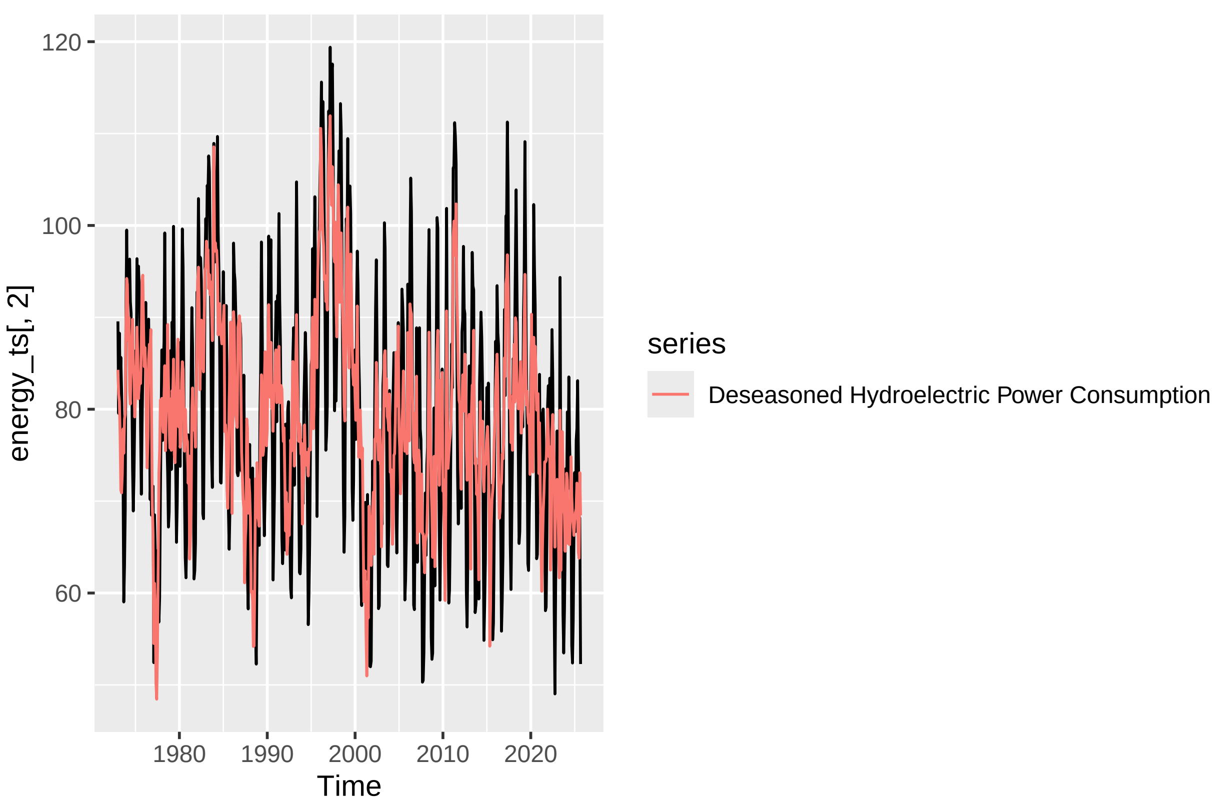
Q8

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?

```r
trp_beta_int <- trp_seas_means_model$coefficients[1]
trp_beta_coeff <- trp_seas_means_model$coefficients[2:12]

hc_beta_int <- hc_seas_means_model$coefficients[1]
hc_beta_coeff <- hc_seas_means_model$coefficients[2:12]

nobs <- nrow(energy_subset)

trp_season_comp <- array(0, nobs)
hc_season_comp <- array(0, nobs)
for (i in 1:nobs) {
  trp_season_comp[i] <- trp_beta_int + trp_beta_coeff%*%trp_dummies[i,]
  hc_season_comp[i] <- hc_beta_int + hc_beta_coeff%*%hc_dummies[i,]
}

trp_deseasoned <- energy_ts[, 1] - trp_season_comp
hc_deseasoned <- energy_ts[, 2] - hc_season_comp

autoplot(energy_ts[, 1]) +
  autolayer(trp_deseasoned, series = "Deseasoned Total Renewable Energy Production")
```




```r
autoplot(energy_ts[, 2]) +
  autolayer(hc_deseasoned, series = "Deseasoned Hydroelectric Power Consumption")
```

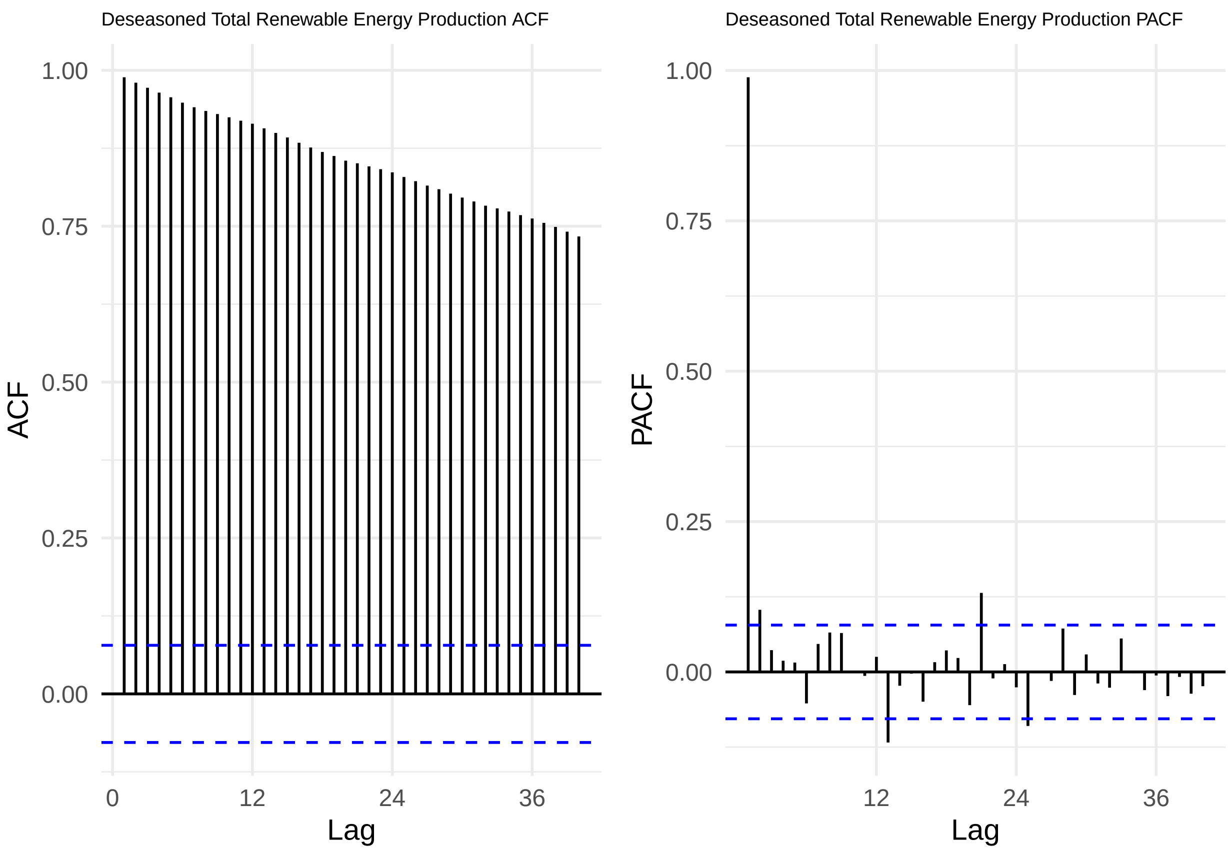


Answer: The overall trends are the same but the short-term peaks and troughs are not as extreme.

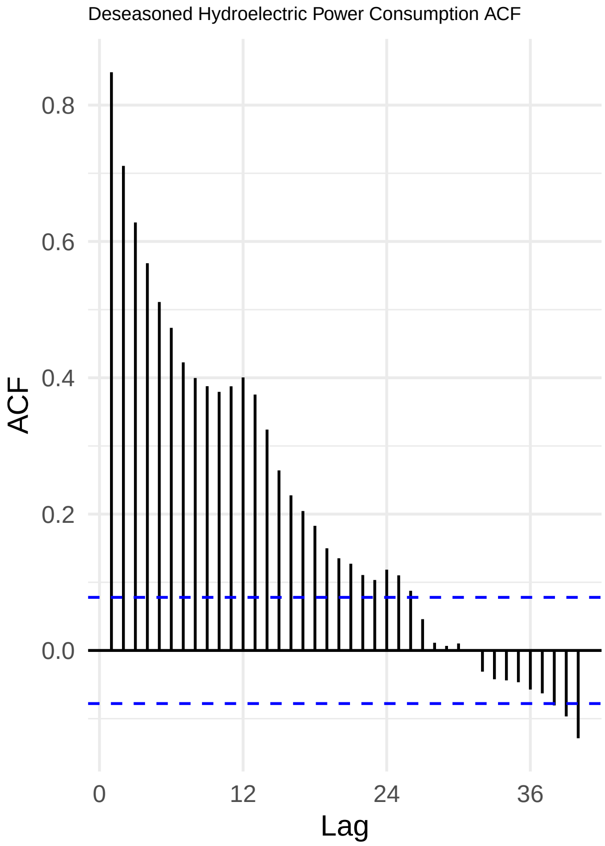
Q9

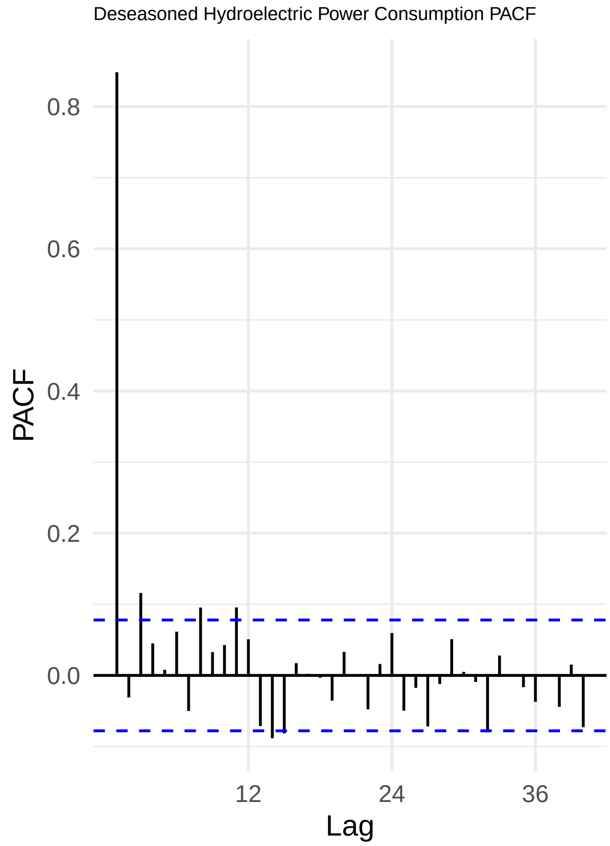
Plot ACF and PACF for the deseason series and compare with the plots from Q1. You may use plot_grid() again to get them side by side. Did the plots change? How?

```
trp_deseasoned_acf <- ggAcf(trp_deseasoned, lag.max = 40, plot = TRUE) +
  ggtitle(paste("Deseasoned Total Renewable Energy Production ACF")) +
  theme_minimal() +
  theme(plot.title = element_text(size = 7))

trp_deseasoned_pacf <- ggPacf(trp_deseasoned, lag.max = 40, plot = TRUE) +
  ggtitle(paste("Deseasoned Total Renewable Energy Production PACF")) +
  theme_minimal() +
  theme(plot.title = element_text(size = 7))

plot_grid(trp_deseasoned_acf, trp_deseasoned_pacf, nrow = 1)
```




```
hc_deseasoned_acf <- ggAcf(hc_deseasoned, lag.max = 40, plot = TRUE) +
  ggtitle(paste("Deseasoned Hydroelectric Power Consumption ACF")) +
  theme_minimal() +
  theme(plot.title = element_text(size = 7))

hc_deseasoned_pacf <- ggPacf(hc_deseasoned, lag.max = 40, plot = TRUE) +
  ggtitle(paste("Deseasoned Hydroelectric Power Consumption PACF")) +
  theme_minimal() +
  theme(plot.title = element_text(size = 7))

plot_grid(hc_deseasoned_acf, hc_deseasoned_pacf, nrow = 1)
```





Answer: The deseasnoned Total Renewable Energy Production ACF looks the same as the original, while the deseasoned PACF has values with absolute values that are lower than those in the original. The deseasoned Hydroelectric Power Consumption ACF has looks to be nearly continually decreasing with some extra component that is roughly a sigmoid, this is in contrast to the original which looks more like a sigmoid with some steadily decreasing component (difference is in the strength of the sigmoid component vs the steadily decreasing component). The deseasoned Hydroelectric Power Consumption PACF has values with absolute values that are lower than those in the original.