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

Assignment 3 - Due date 02/03/26

Josh Salzberg

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("tidyverse")
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.6
## v forcats   1.0.1     v stringr   1.6.0
```

```
## v ggplot2   4.0.1     v tibble    3.3.1
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.2.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library("here")
```

```
## here() starts at C:/Users/jhsal/OneDrive - Duke University/797/TSA2026/TSA_Sp26
```

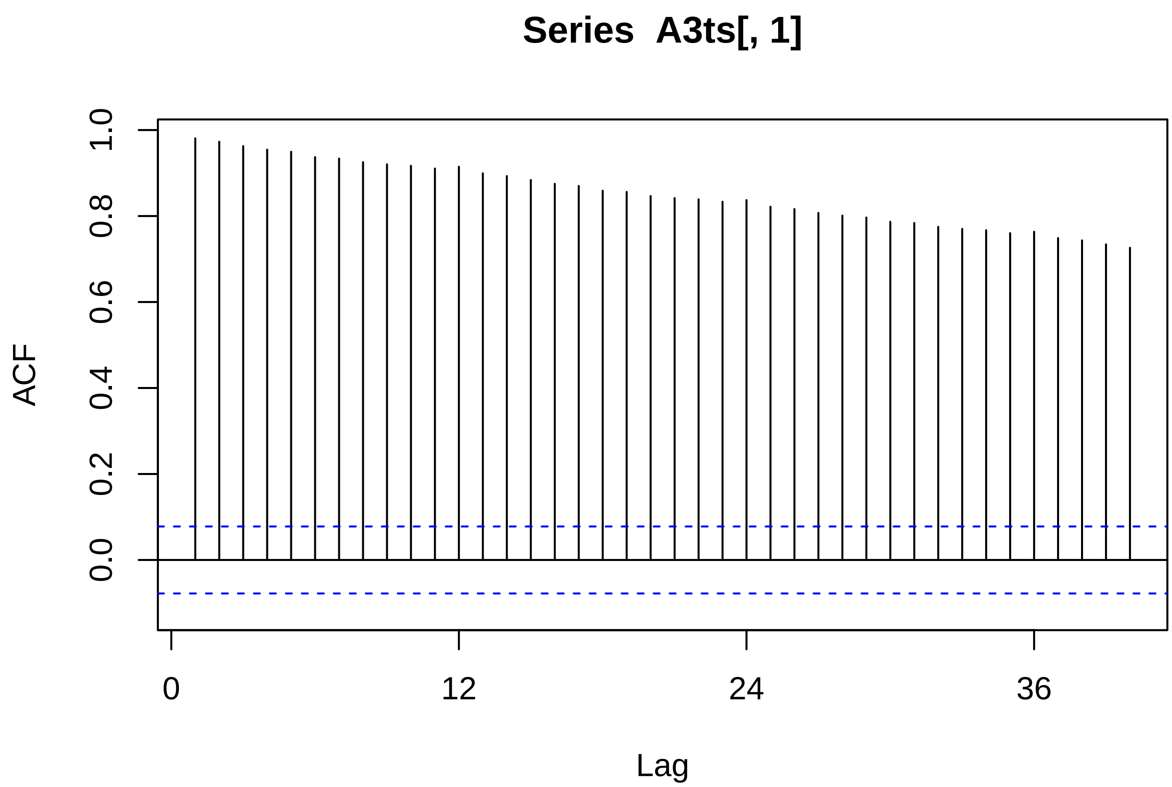
```
# Reading in energy data
energydata <- read.csv(here("Data", "Processed","jhs_clean_energy_data.csv"))
# Selecting energy data
A3data <- energydata %>%
  select(TotalRenewables = Total.Renewable.Energy.Production,
         HydroelectricConsumption = Hydroelectric.Power.Consumption)
# convert to time series object
A3ts <- ts(A3data, start = c(1973, 1), frequency = 12)
```

##Trend Component

Q1

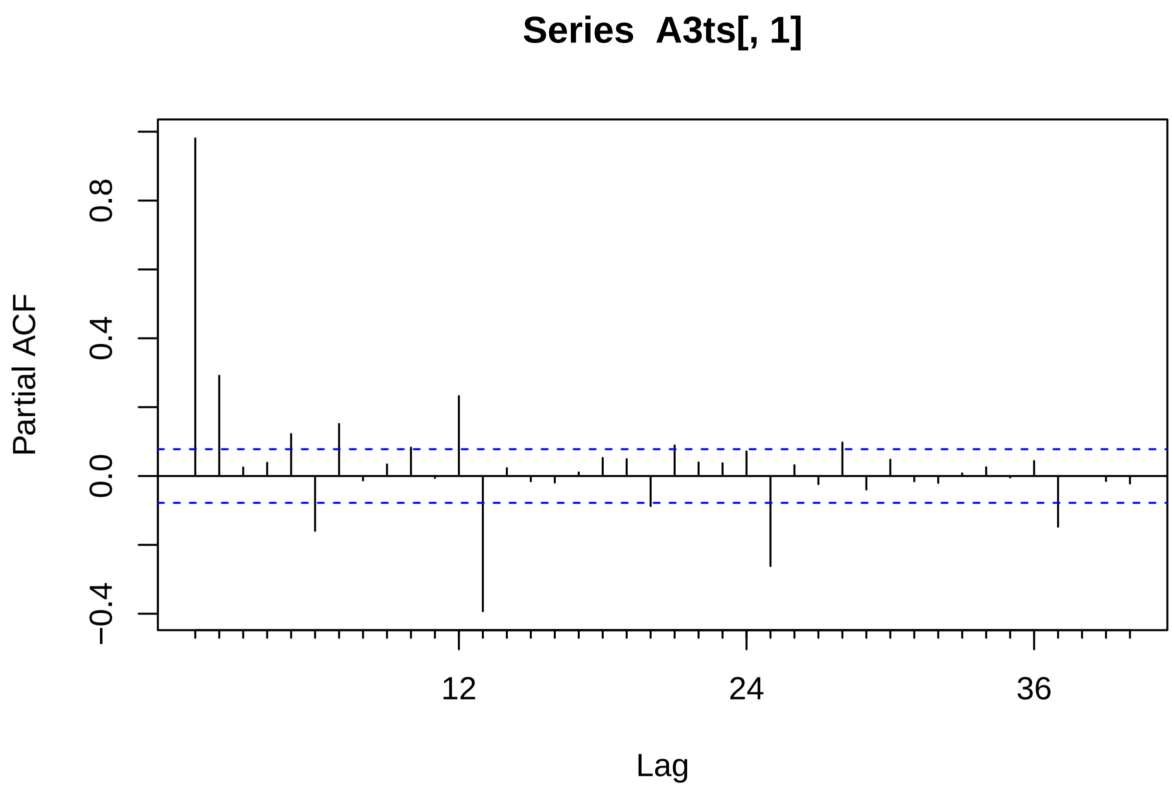
For each series (Total Renewable Production and Hydroelectric Consumption) create three plots arranged in a row (side-by-side): (1) time series plot, (2) ACF, (3) PACF. Use cowplot::plot_grid() to place them in a grid.

```
Renewables1 <- autoplot(A3ts[,1],
         main ="Time Series",
         x = "Date (Monthly Data)",
         y ="Trillion Btu")
Renewables2 <- autoplot(Acf(A3ts[,1], lag.max = 40), main = "ACF")
```



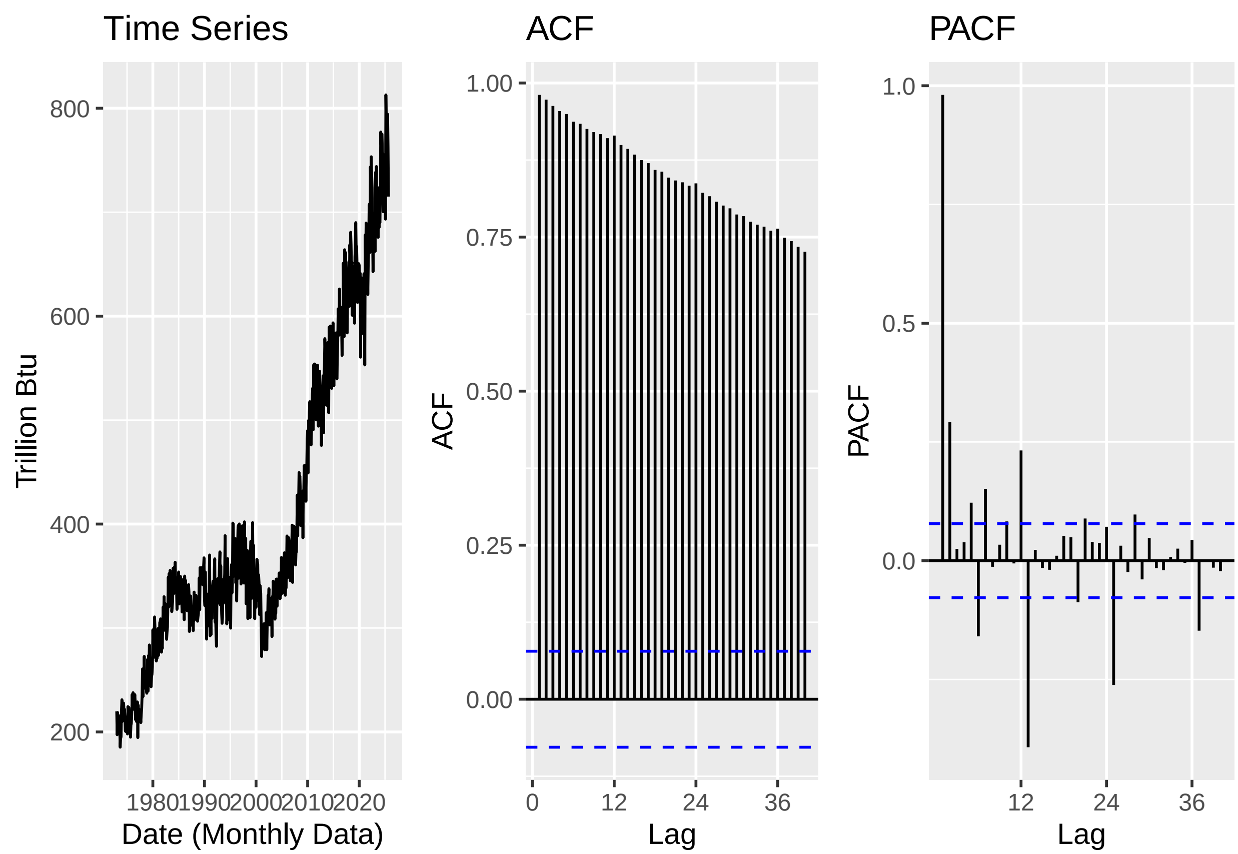

```
## Warning in ggplot2::geom_segment(lineend = "butt", ...): Ignoring unknown
## parameters: `main`
```

```
Renewables3 <- autoplot(Pacf(A3ts[,1], lag.max = 40), main = "PACF")
```

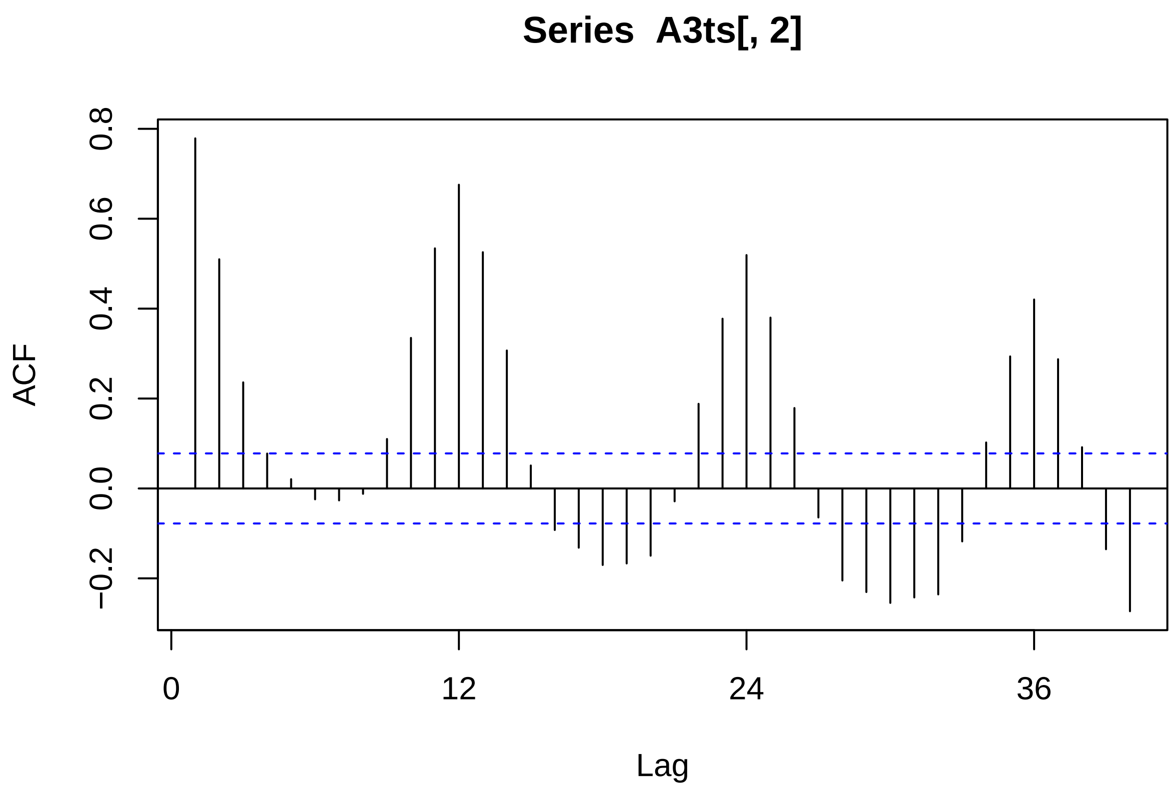



```
## Warning in ggplot2::geom_segment(lineend = "butt", ...): Ignoring unknown
## parameters: `main`
```

```
cowplot::plot_grid(Renewables1, Renewables2, Renewables3, nrow = 1)
```

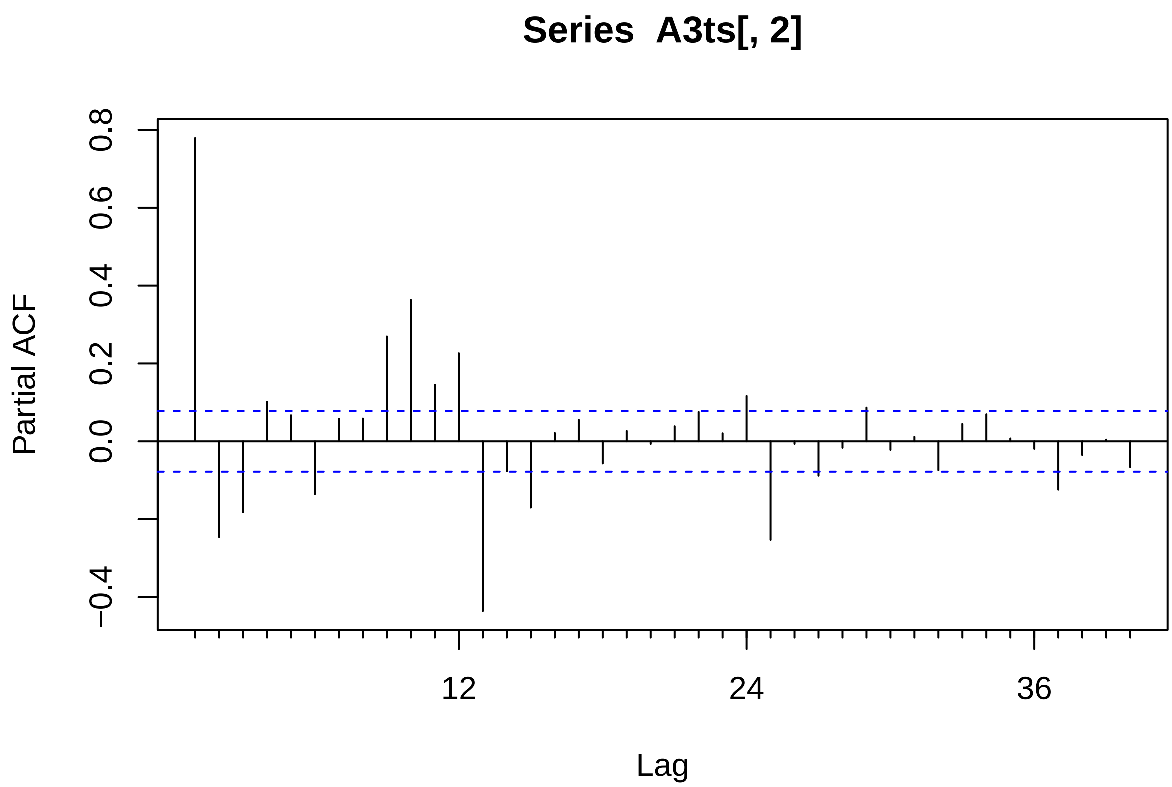



```
Hydro1 <- autoplot(A3ts[,2],
         main ="Time Series",
         x = "Date (Monthly Data)",
         y ="Trillion Btu")
Hydro2 <- autoplot(Acf(A3ts[,2], lag.max = 40), main = "ACF")
```



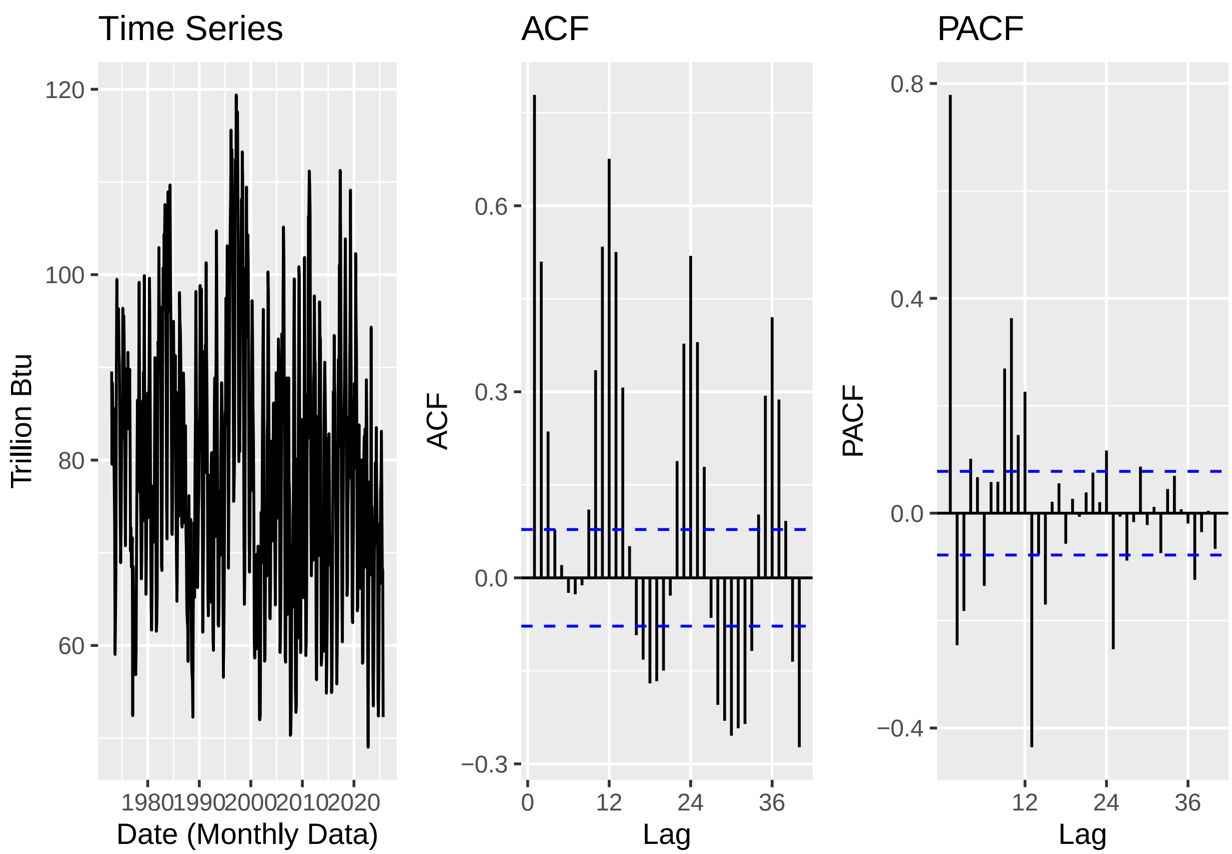

```
## Warning in ggplot2::geom_segment(lineend = "butt", ...): Ignoring unknown
## parameters: `main`
```

```
Hydro3 <- autoplot(Pacf(A3ts[,2], lag.max = 40), main = "PACF")
```




```
## Warning in ggplot2::geom_segment(lineend = "butt", ...): Ignoring unknown
## parameters: `main`
```

```
cowplot::plot_grid(Hydro1, Hydro2, Hydro3, 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 Renewabble Energy Production appears to have a positive trend. Hydroelectric Power Consumption is difficult to ascertain; there may be a slight negative trend towards the end of the time series.

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.

```
#Vector for linear regression
t <- c(1:nrow(A3data))

# renewables regression and coefficients
LM1 <- lm(A3data[,1]~t )
summary(LM1)
```

```
##
## Call:
## lm(formula = A3data[, 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) 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
```

```
LM1Beta0 <- as.numeric(LM1$coefficients[1]) # Intercept
LM1Beta1 <- as.numeric(LM1$coefficients[2]) # Slope

# Hydroelectric regression and coefficients
LM2 <- lm(A3data[,2]~t )
summary(LM2)
```

```
##
## Call:
## lm(formula = A3data[, 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) 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
```

```
LM2Beta0 <- as.numeric(LM2$coefficients[1]) # Intercept
LM2Beta1 <- as.numeric(LM2$coefficients[2]) # Slope
```

Q4

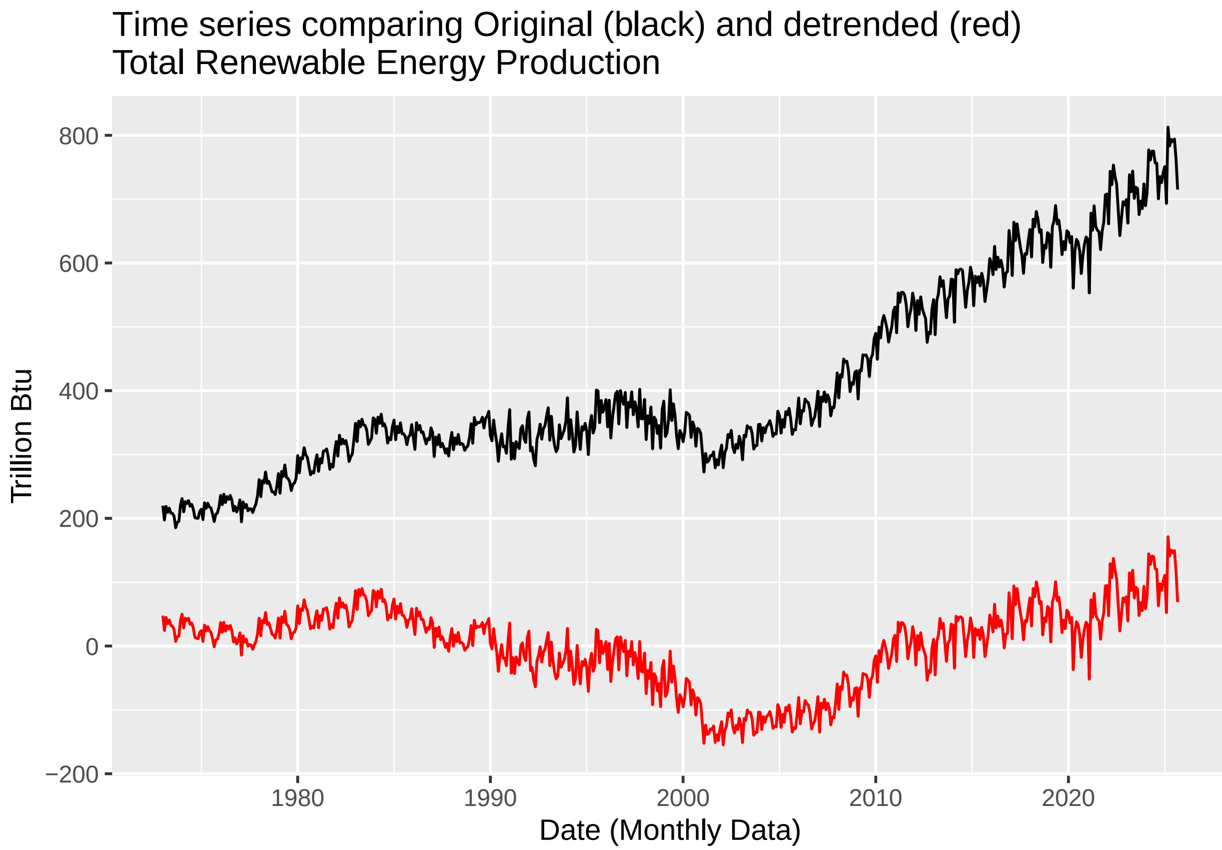
Use the regression coefficients to detrend each series (subtract fitted linear trend). Plot detrended series and compare with the original time series from Q1. Describe what changed.

```
# Detrending Renewable Data
detrend_renewables <- A3data[,1]-(LM1Beta0+LM1Beta1*t)
class(detrend_renewables)
```

```
## [1] "numeric"
```

```
ts_detrend_renewables <- ts(detrend_renewables, frequency=12, start=c(1973,1))
autoplot(A3ts[,1])+
  autolayer(ts_detrend_renewables, color = "red")+
  labs(title = "Time series comparing Original (black) and detrended (red) \nTotal Renewable Energy Pro
```

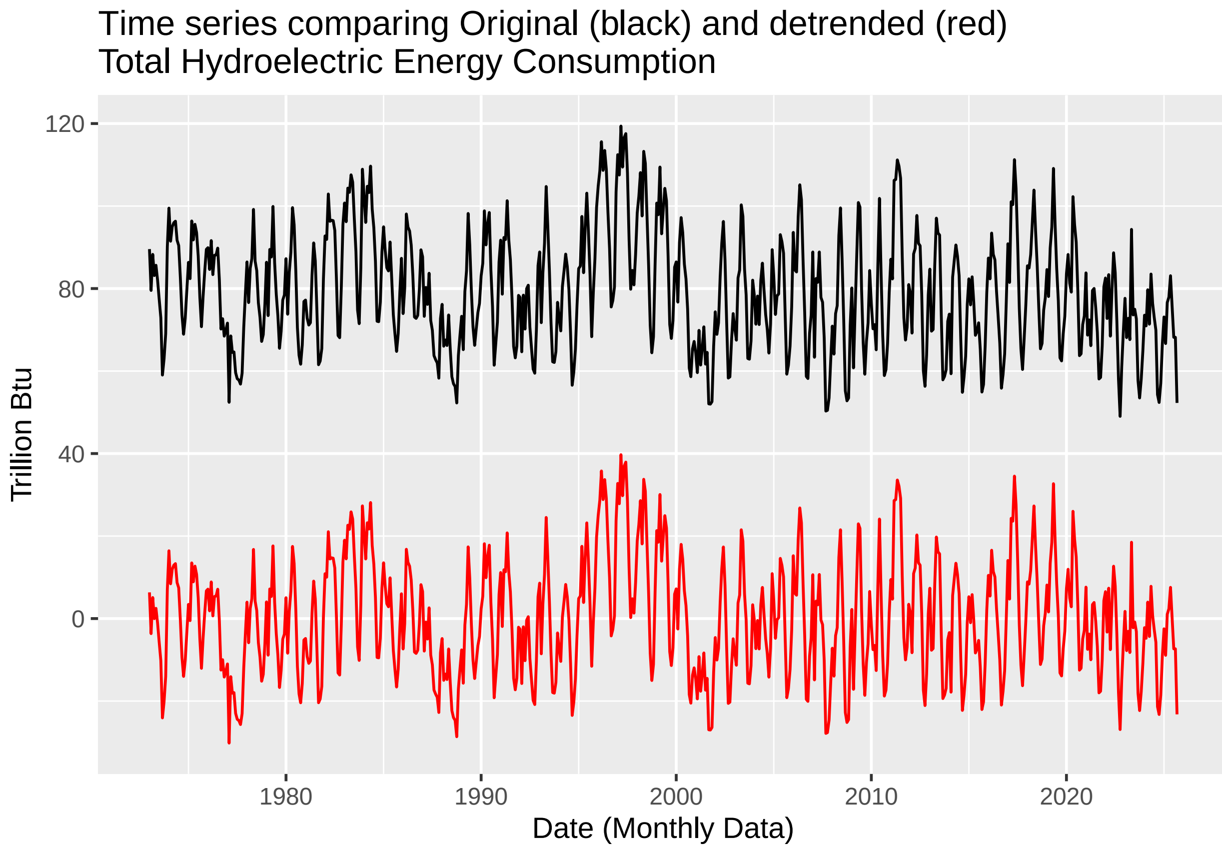
```
       x = "Date (Monthly Data)",
       y ="Trillion Btu")
```




```
# Detrending Hydroelectric Data
detrend_hydro <- A3data[,2]-(LM2Beta0+LM2Beta1*t)
class(detrend_hydro)
```

```
## [1] "numeric"
```

```
ts_detrend_hydro <- ts(detrend_hydro, frequency=12, start=c(1973,1))
autoplot(A3ts[,2])+
  autolayer(ts_detrend_hydro, color = "red")+
  labs(title = "Time series comparing Original (black) and detrended (red) \nTotal Hydroelectric Energy
       x = "Date (Monthly Data)",
       y ="Trillion Btu")
```

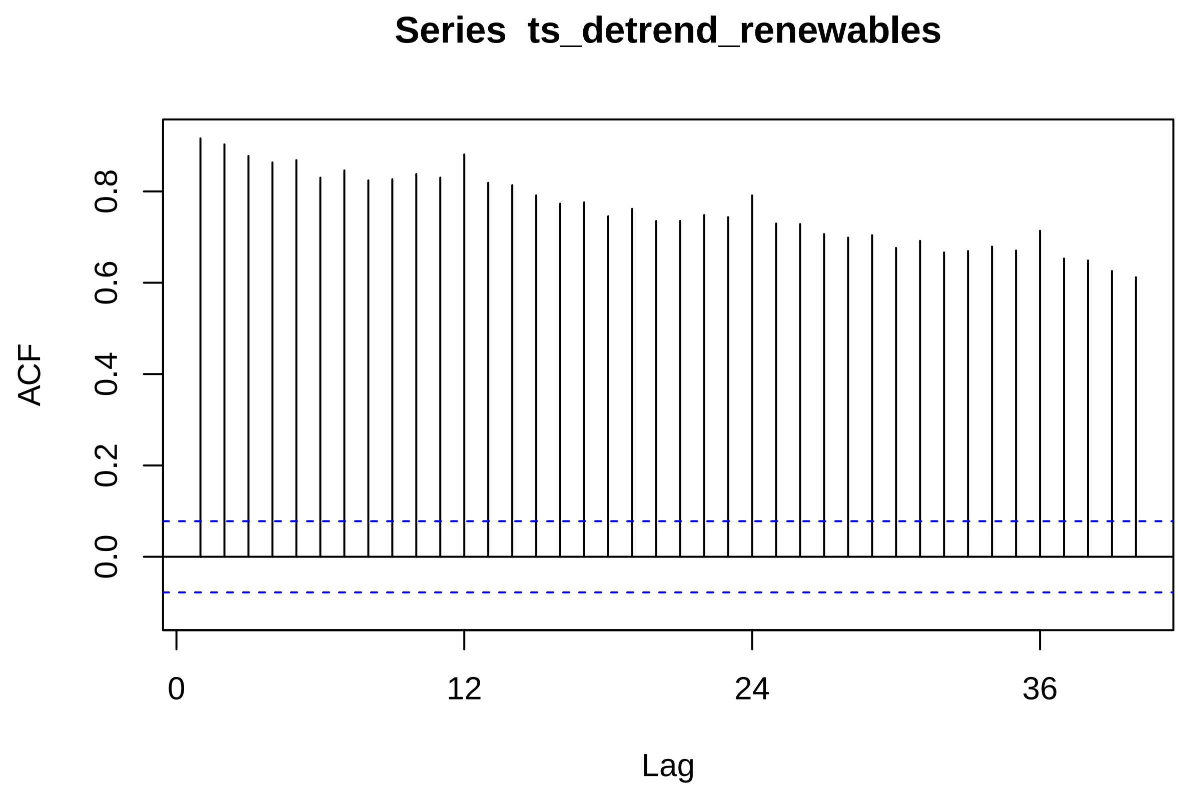


Answer: The detrended datasets are now centered along the y = 0 line. The shape of the detrended renewables chart is much flatter now than previously; the overall shape of the hydroelectric data has not changed because there was very little trend in the data.

Q5

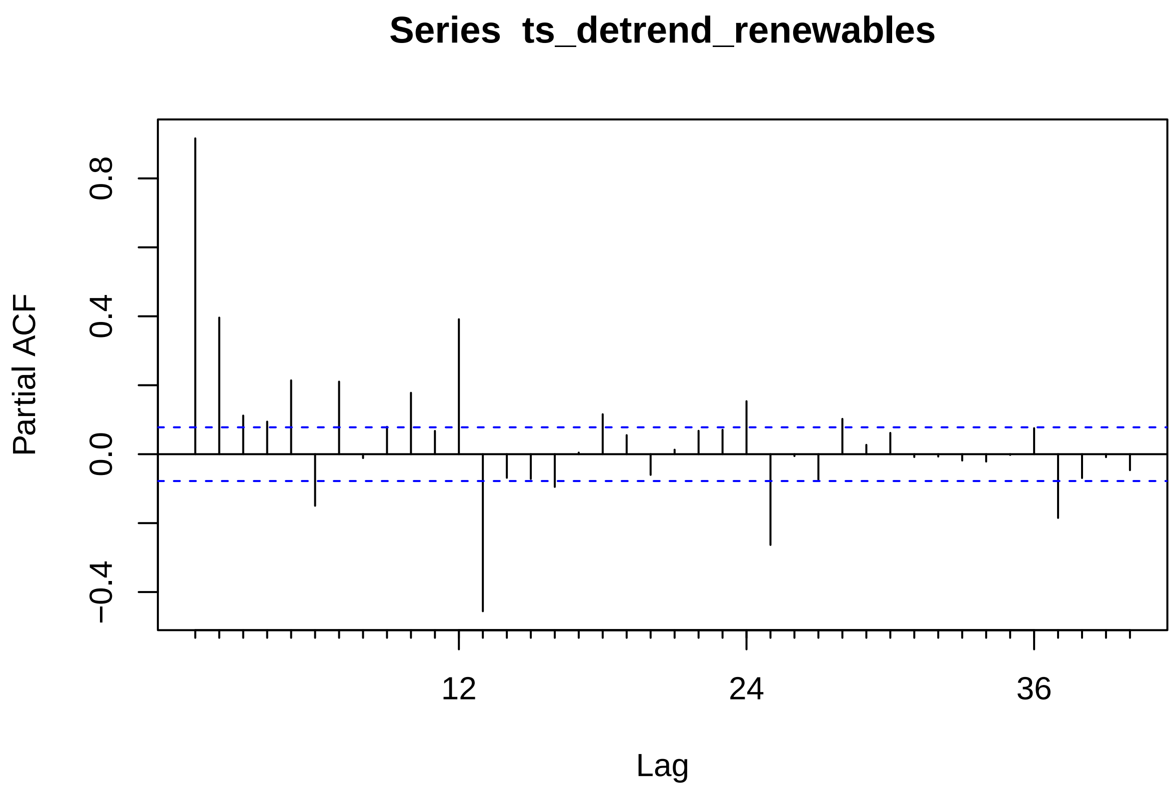
Plot ACF and PACF for the detrended series and compare with the plots from Q1. You may use plot_grid() again to get them side by side to make it easier to compare. Did the plots change? How?

```
# Comparing Renewables ACF & PACF from Original and Detrend
RenewablesDetrendAcf <- autoplot(Acf(ts_detrend_renewables, lag.max = 40), main = "Detrend ACF")
```



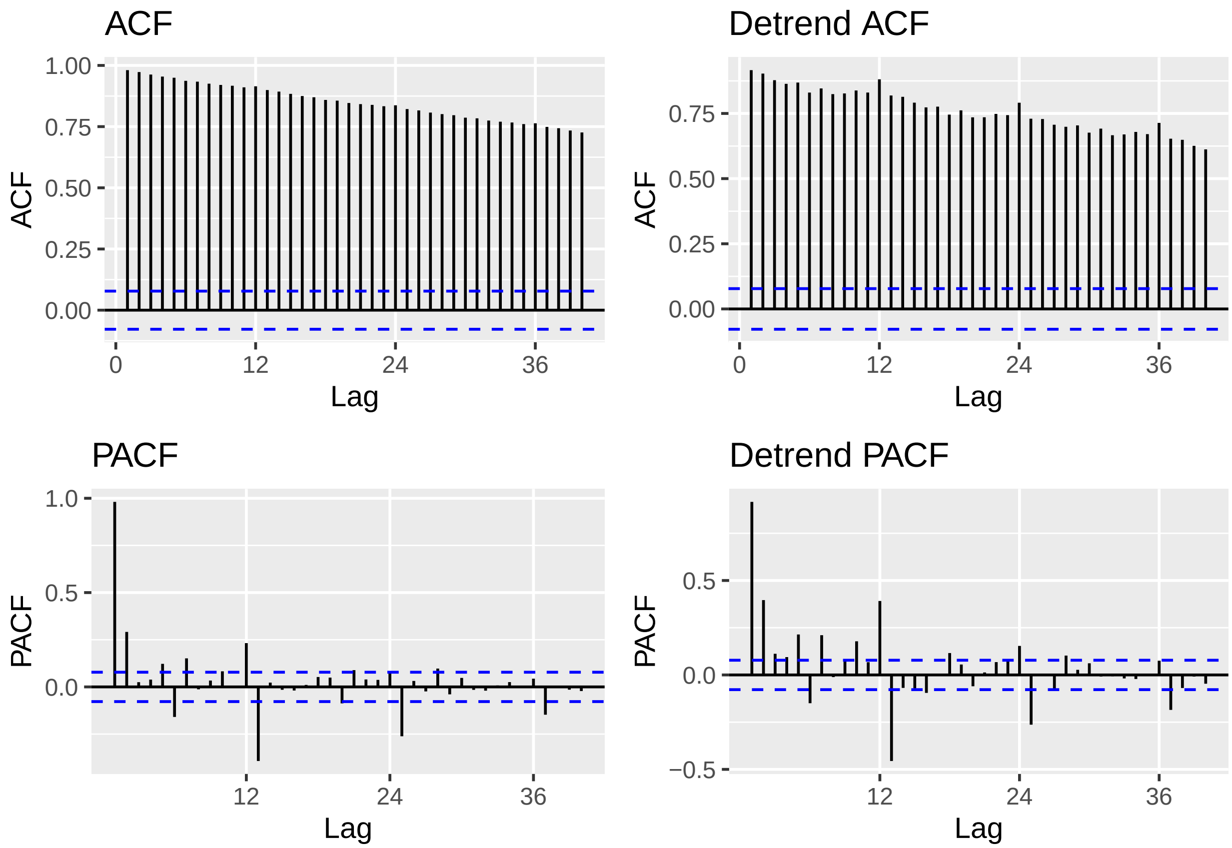

```
## Warning in ggplot2::geom_segment(lineend = "butt", ...): Ignoring unknown
## parameters: `main`
```

```r
RenewablesDetrendPacf <- autoplot(Pacf(ts_detrend_renewables, lag.max = 40), main = "Detrend PACF")
```

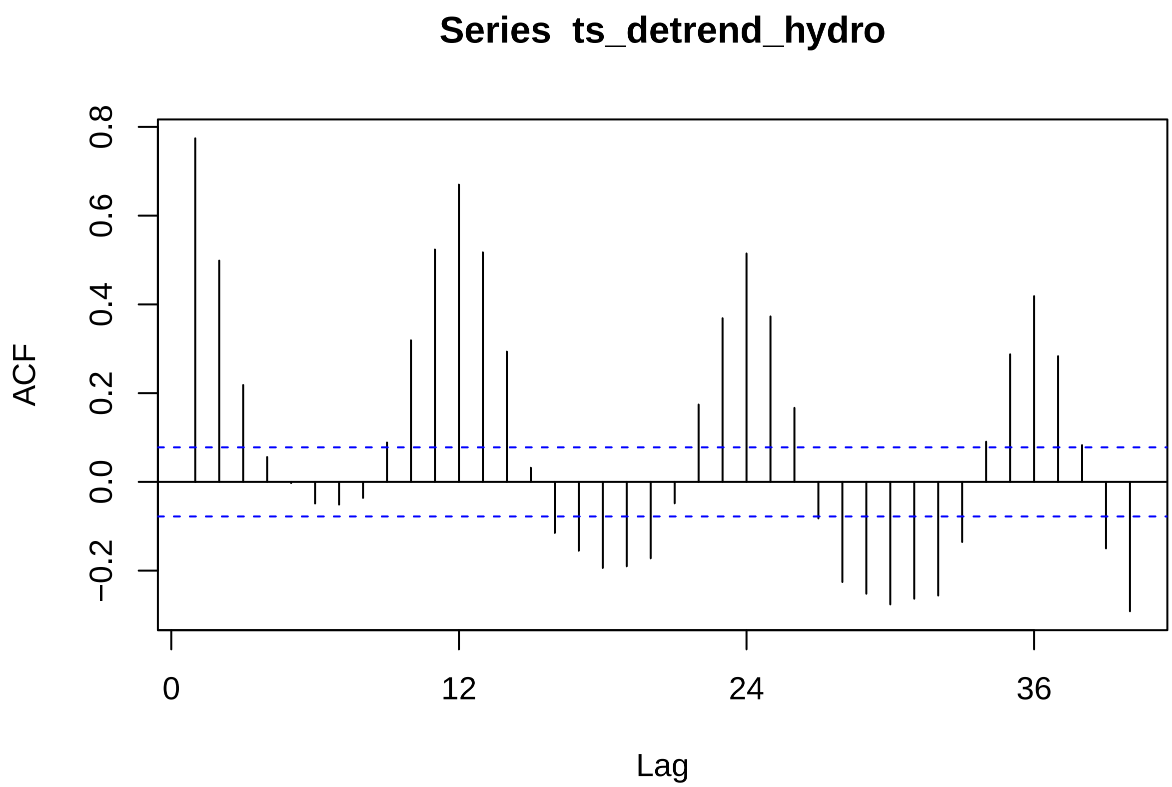



```
## Warning in ggplot2::geom_segment(lineend = "butt", ...): Ignoring unknown
## parameters: `main`
```

```
cowplot::plot_grid(Renewables2, RenewablesDetrendAcf,
          Renewables3, RenewablesDetrendPacf,
          nrow = 2, ncol = 2)
```

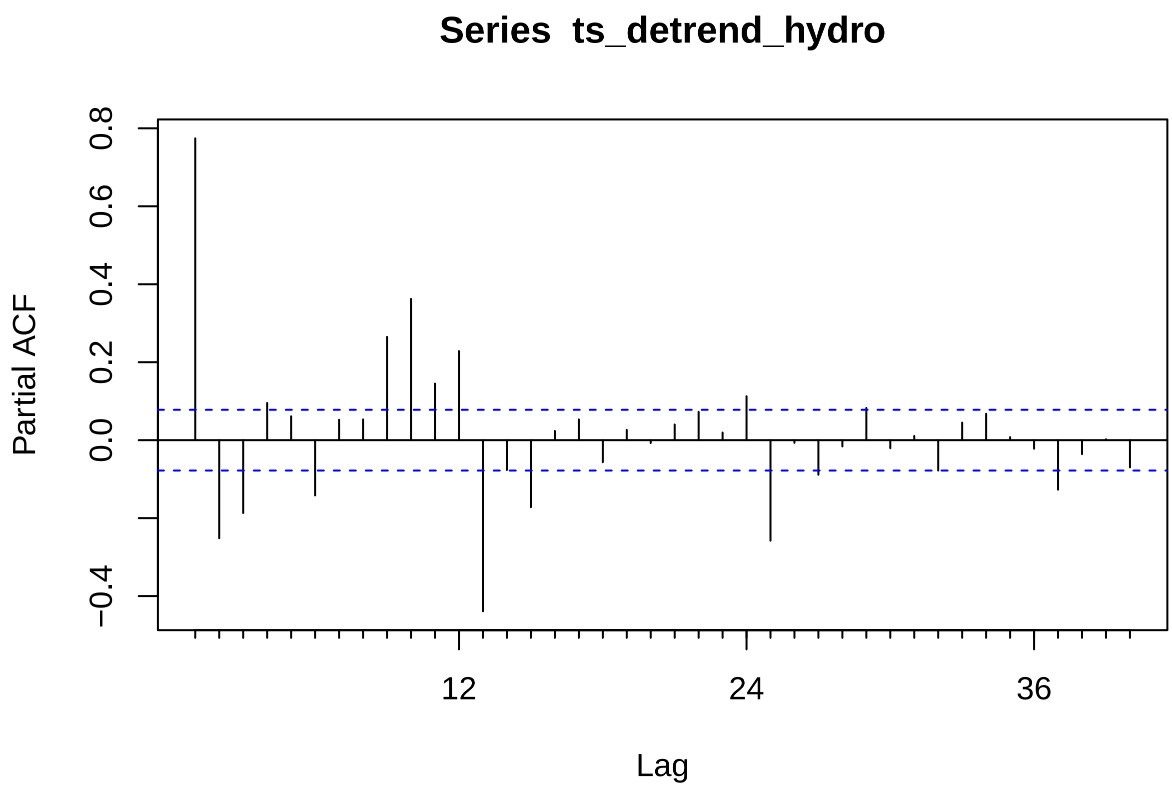



```r
# Comparing Hydroelectric ACF & PACF from Original and Detrend
HydroDetrendAcf <- autoplot(Acf(ts_detrend_hydro, lag.max = 40), main = "Detrend ACF")
```



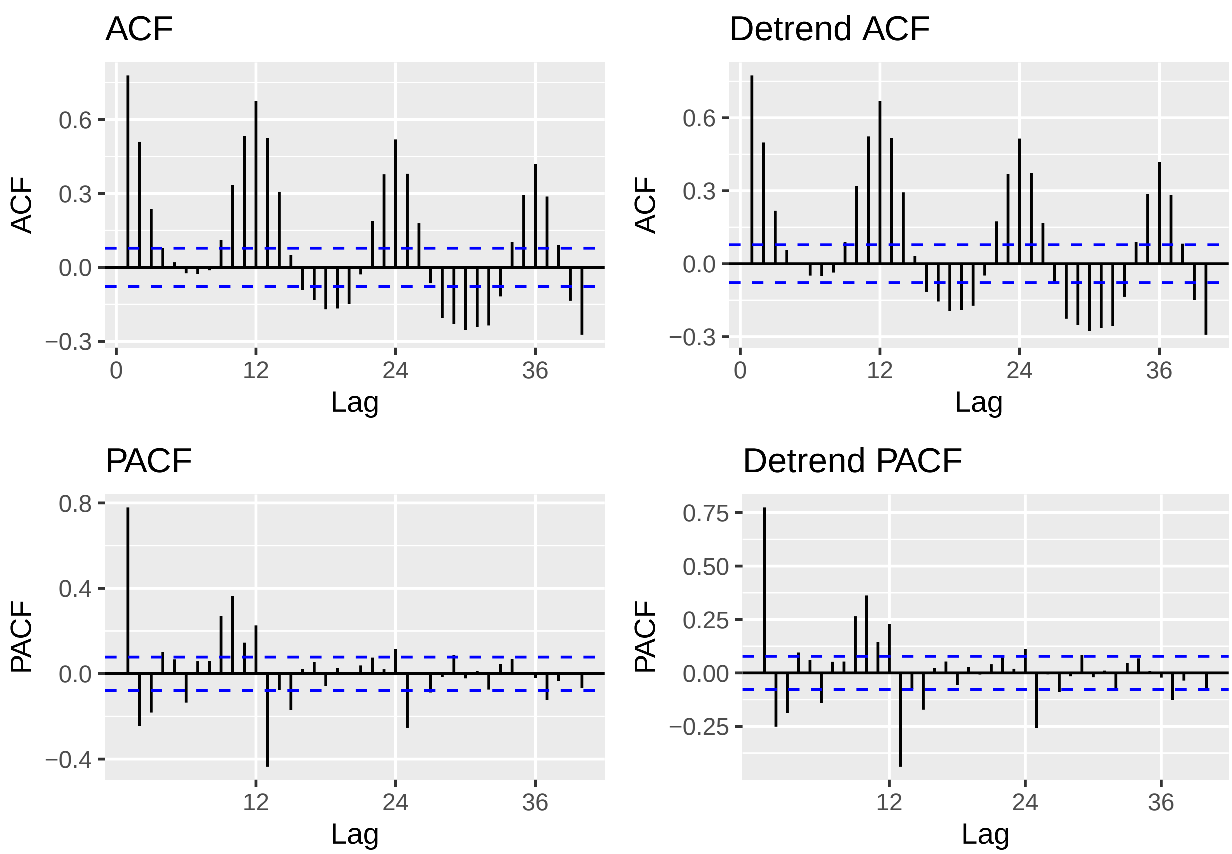

```
## Warning in ggplot2::geom_segment(lineend = "butt", ...): Ignoring unknown
## parameters: `main`
```

```
HydroDetrendPacf <- autoplot(Pacf(ts_detrend_hydro, lag.max = 40), main = "Detrend PACF")
```




```
## Warning in ggplot2::geom_segment(lineend = "butt", ...): Ignoring unknown
## parameters: `main`
```

```
cowplot::plot_grid(Hydro2, HydroDetrendAcf,
          Hydro3, HydroDetrendPacf,
          nrow = 2, ncol = 2)
```



answer: the Renewables datagraphs change notably; most significantly, the detrended data are less significantly correlated, demonstrated by the shifted scales on the Y-axis. Additionally, the both detrended charts demonstrate more significant seasonality. On the other hand, the hydroelectric data changed very little, reflecting that trend generally plays a small role in the time variation.

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: The Total Renewables data appears to demonstrate fairly regular but limited impact seasonality. The Hydroelectric data demonstrates very strong seasonality with a large magnitude of effect.

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?

```
# prime the dummies
dummies <- seasonaldummy(A3ts[,1])
nrow(dummies)
```

```
## [1] 633
```

```r
#regress each series
#renewables
seasonalRenewables <- lm(detrend_renewables~dummies)
summary(seasonalRenewables)
```

```
##
## Call:
## lm(formula = detrend_renewables ~ 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
## dummiesJan     5.840     12.334   0.473   0.6360
## dummiesFeb   -31.525     12.334  -2.556   0.0108 *
## dummiesMar     8.205     12.334   0.665   0.5061
## dummiesApr    -5.400     12.334  -0.438   0.6617
## dummiesMay     8.912     12.334   0.723   0.4703
## dummiesJun    -2.231     12.334  -0.181   0.8565
## dummiesJul     3.114     12.334   0.252   0.8008
## dummiesAug    -5.478     12.334  -0.444   0.6571
## dummiesSep   -31.283     12.334  -2.536   0.0114 *
## dummiesOct   -18.437     12.393  -1.488   0.1373
## 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
#hydro
seasonalHydro <- lm(detrend_hydro~dummies)
summary(seasonalHydro)
```

```
##
## Call:
## lm(formula = detrend_hydro ~ 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
## dummiesJan    4.8900     1.9747   2.476 0.013541 *
## dummiesFeb   -2.4641     1.9747  -1.248 0.212573
## dummiesMar    7.0794     1.9747   3.585 0.000364 ***
## dummiesApr    5.5895     1.9747   2.830 0.004798 **
## dummiesMay   14.0676     1.9747   7.124 2.92e-12 ***
```

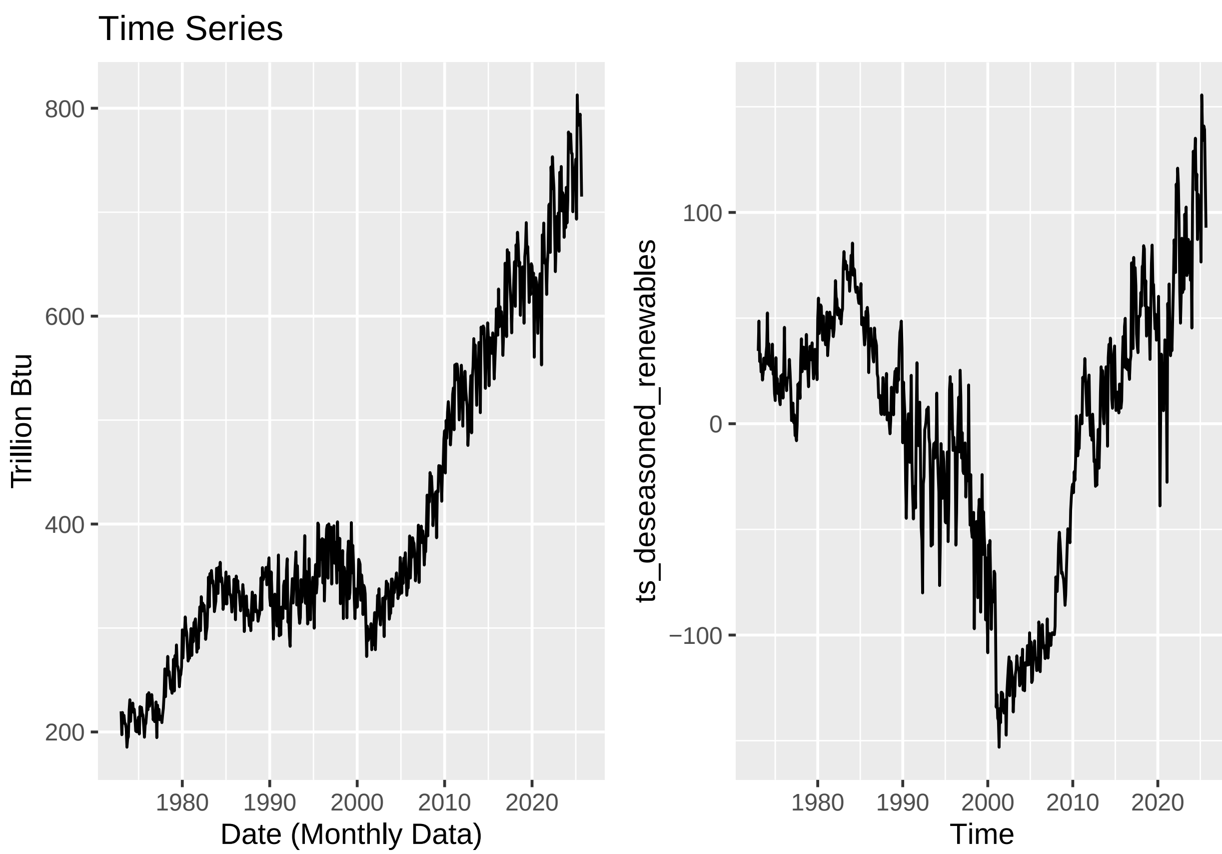
```
## dummiesJun   10.7799     1.9747   5.459 6.93e-08 ***
## dummiesJul    4.0156     1.9747   2.033 0.042427 *
## dummiesAug   -5.2952     1.9747  -2.681 0.007525 **
## dummiesSep  -16.5612     1.9747  -8.386 3.37e-16 ***
## dummiesOct  -16.3534     1.9841  -8.242 1.01e-15 ***
## dummiesNov  -10.7940     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: While the regression of Renewable Production reveals that two months imply somewhat significant seasonality, overall the model is extremely weak, with monthly seasonality explaining roughly 3% of the data. The hydroelectric data, on the other hand, reveals strong seasonality with 11 of 12 months demonstrating statistically significant correlation. Importantly, the model has a moderately strong explanatory impact, demonstrating 47% R^2 metric.

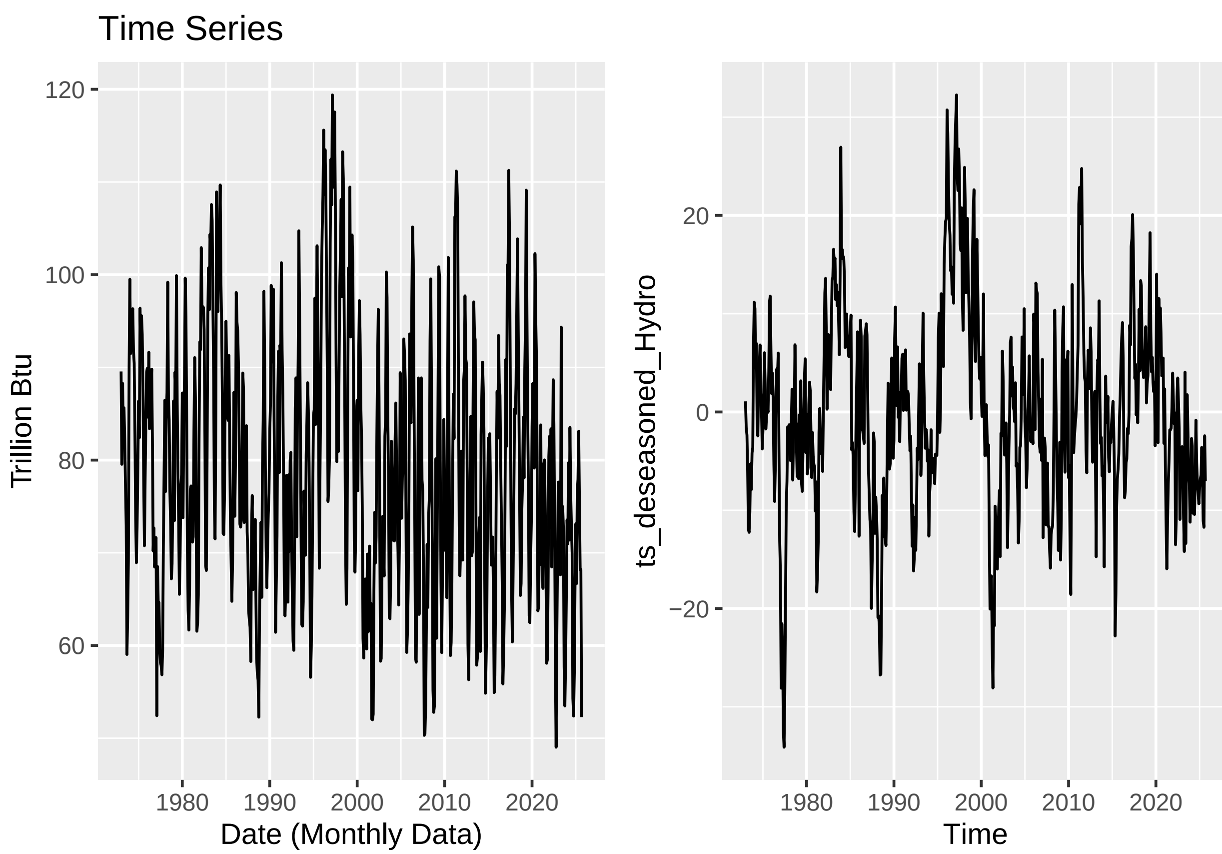
Q8

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 deseasoning
renewable_beta_int <- seasonalRenewables$coefficients[1]
renewable_beta_coeff <- seasonalRenewables$coefficients[2:12]
#Compute the seasonal component
season_renewables_data <- array(0,nrow(A3data))
for(i in t){
  season_renewables_data[i] <- renewable_beta_int+renewable_beta_coeff%*%dummies[i,]
}

# removing seasonal component from series
deseasoned_renewables <- detrend_renewables - season_renewables_data
#convert to time series
ts_deseasoned_renewables <- ts(deseasoned_renewables, start = c(1973, 1), frequency = 12)

#plot
cowplot::plot_grid(Renewables1, autoplot(ts_deseasoned_renewables))
```




```r
# Hydro deseasoning
Hydro_beta_int <- seasonalHydro$coefficients[1]
Hydro_beta_coeff <- seasonalHydro$coefficients[2:12]
#Compute the seasonal component
season_Hydro_data <- array(0,nrow(A3data))
for(i in t){
  season_Hydro_data[i] <- Hydro_beta_int+Hydro_beta_coeff%*%dummies[i,]
}

# removing seasonal component from series
deseasoned_Hydro <- detrend_hydro - season_Hydro_data
#convert to time series
ts_deseasoned_Hydro <- ts(deseasoned_Hydro, start = c(1973, 1), frequency = 12)

#plot
cowplot::plot_grid(Hydro1, autoplot(ts_deseasoned_Hydro))
```

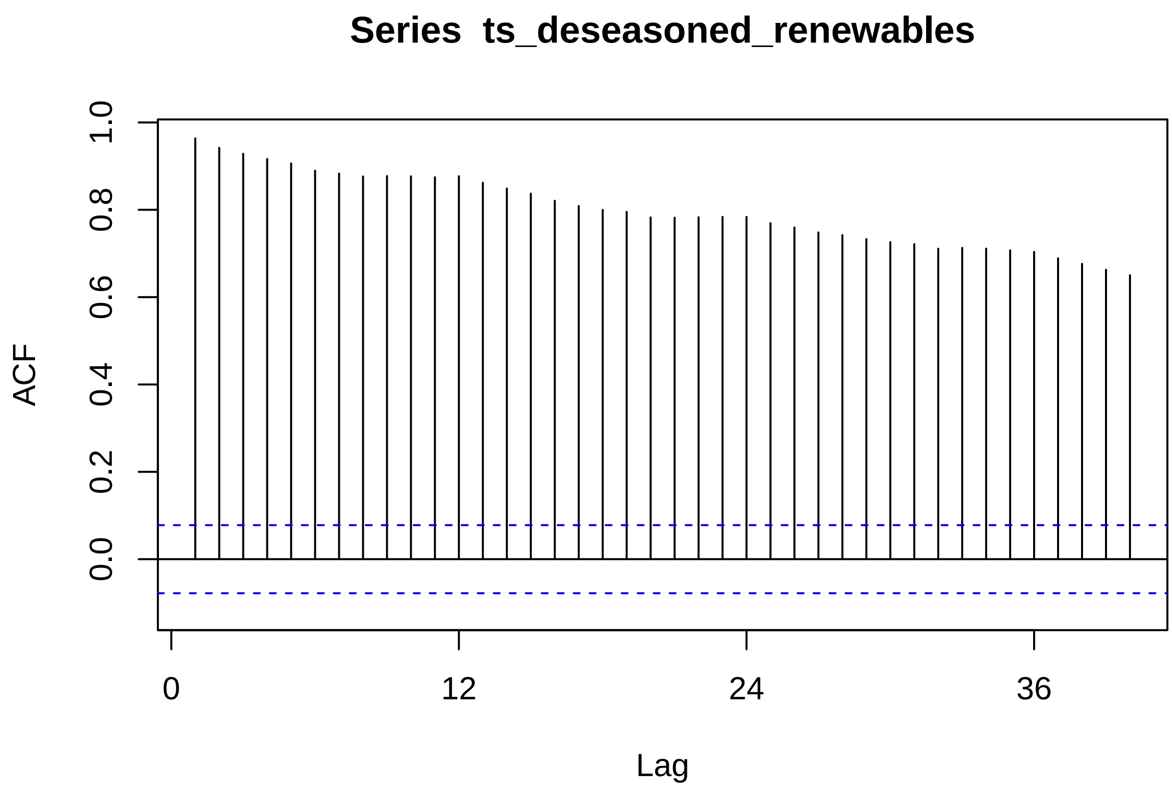


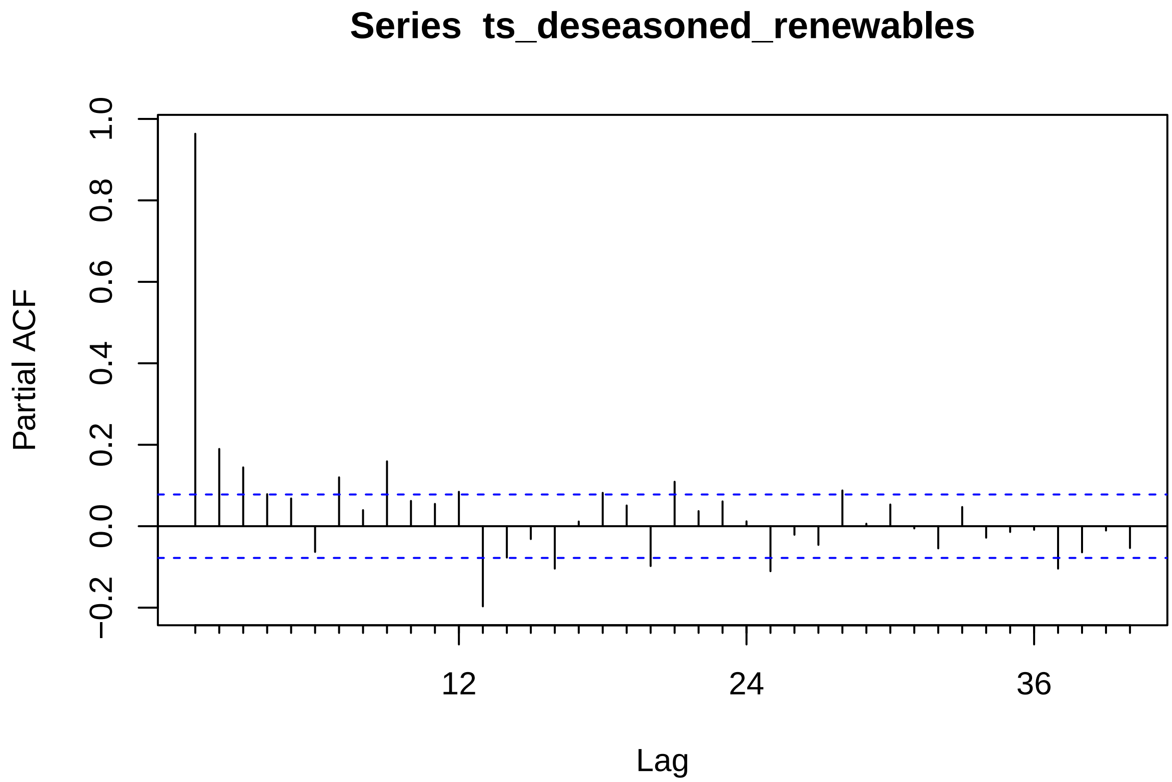
Answer: Much has changed! The Renewables time series has both flattened around the Y=0 line as well as lst some of the minute seasonal changes. The Hydroelectric time series demonstrates a much smoother, less jittery chart. We can almost detect curves to the data rather than erratic oscillations.

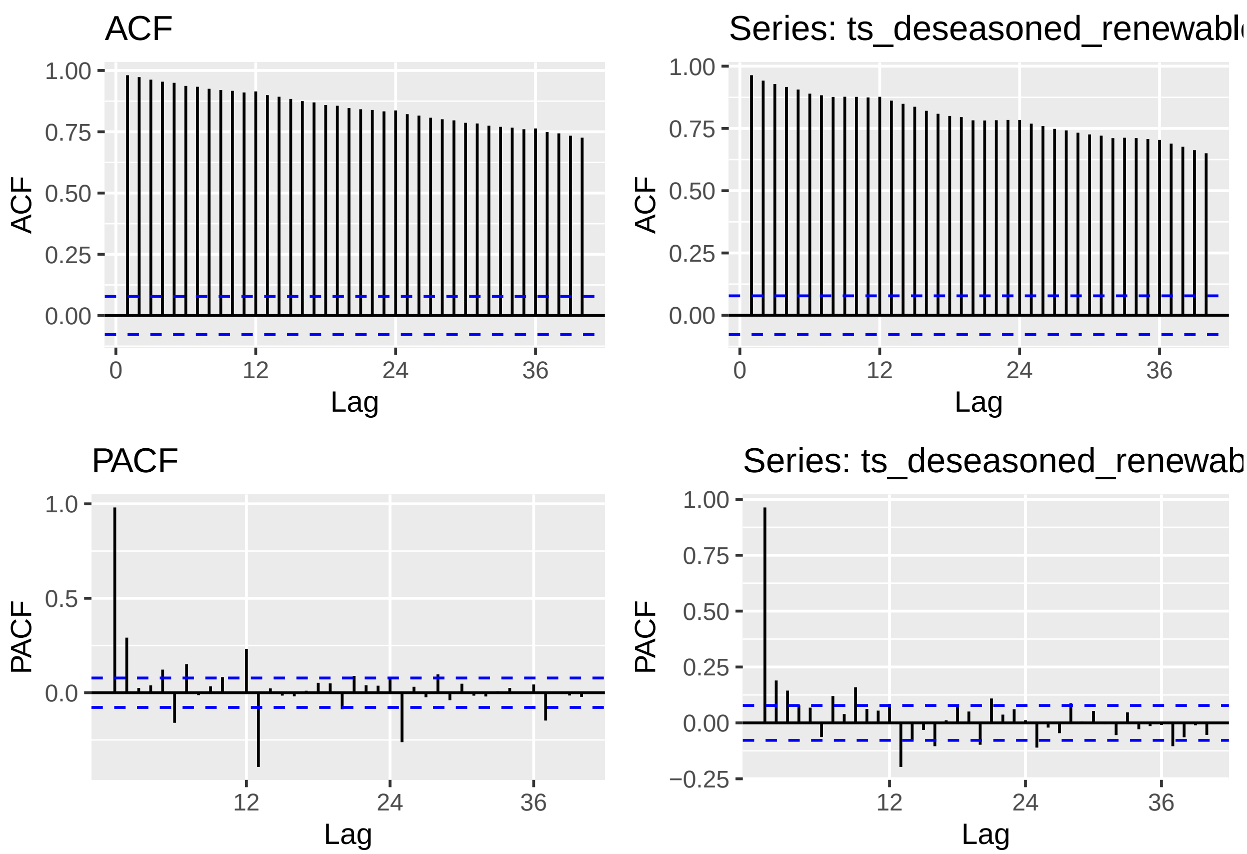
Q9

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?

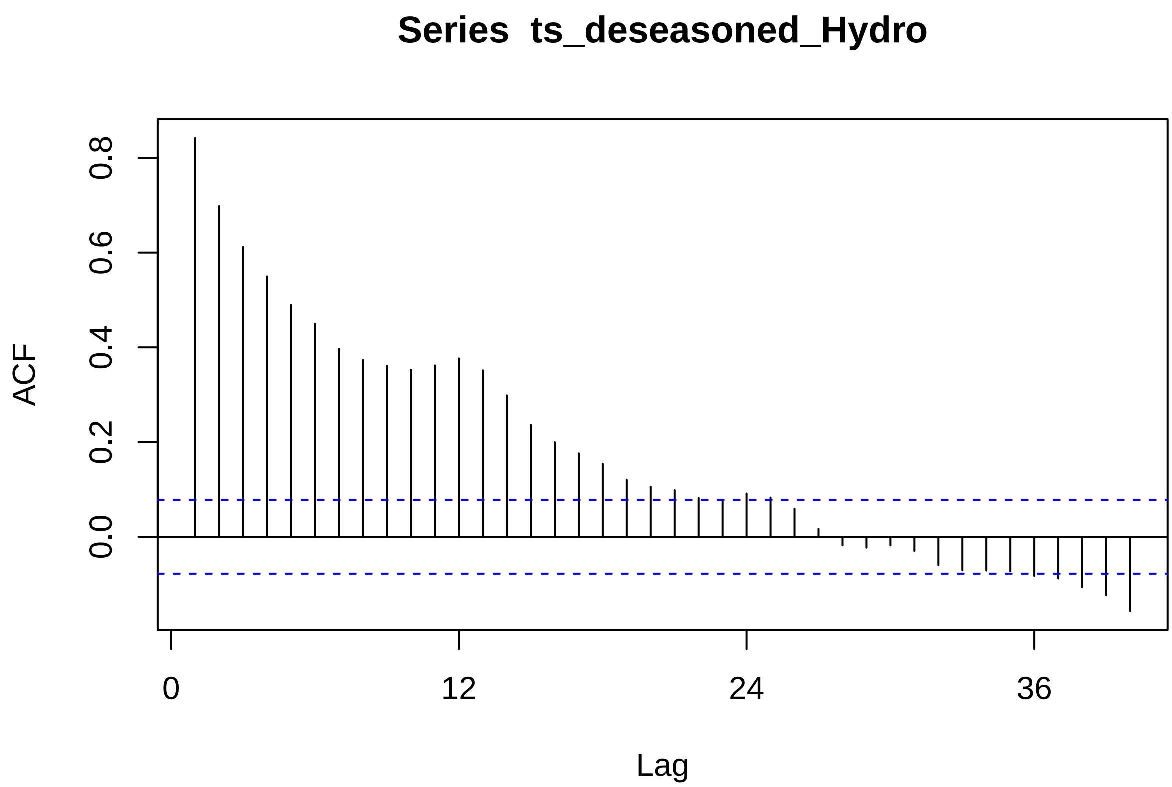
```
cowplot::plot_grid(Renewables2, autoplot(Acf(ts_deseasoned_renewables, lag.max = 40)),
                   Renewables3, autoplot(Pacf(ts_deseasoned_renewables, lag.max = 40)),
                   nrow = 2, ncol = 2)
```

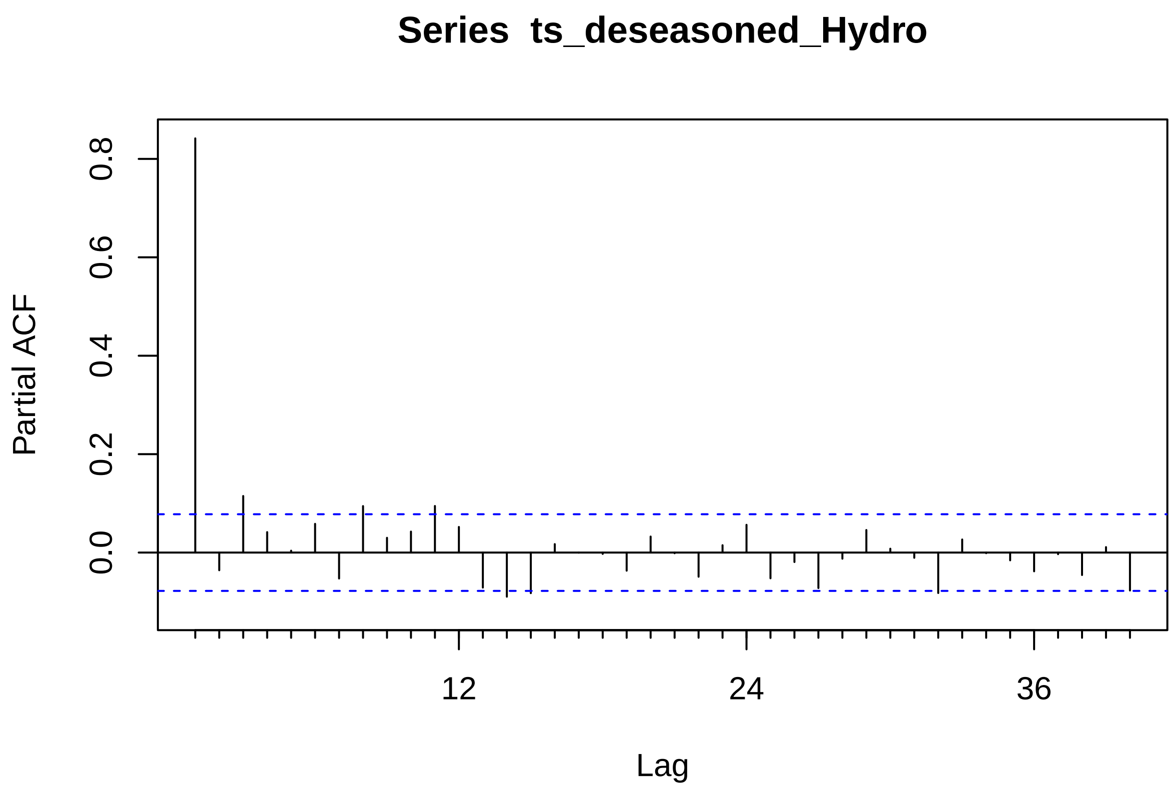

Series ts_deseasoned_renewables
ACF
1.0
0.8
0.6
0.4
0.2
0.0
0
12
24
36
Lag

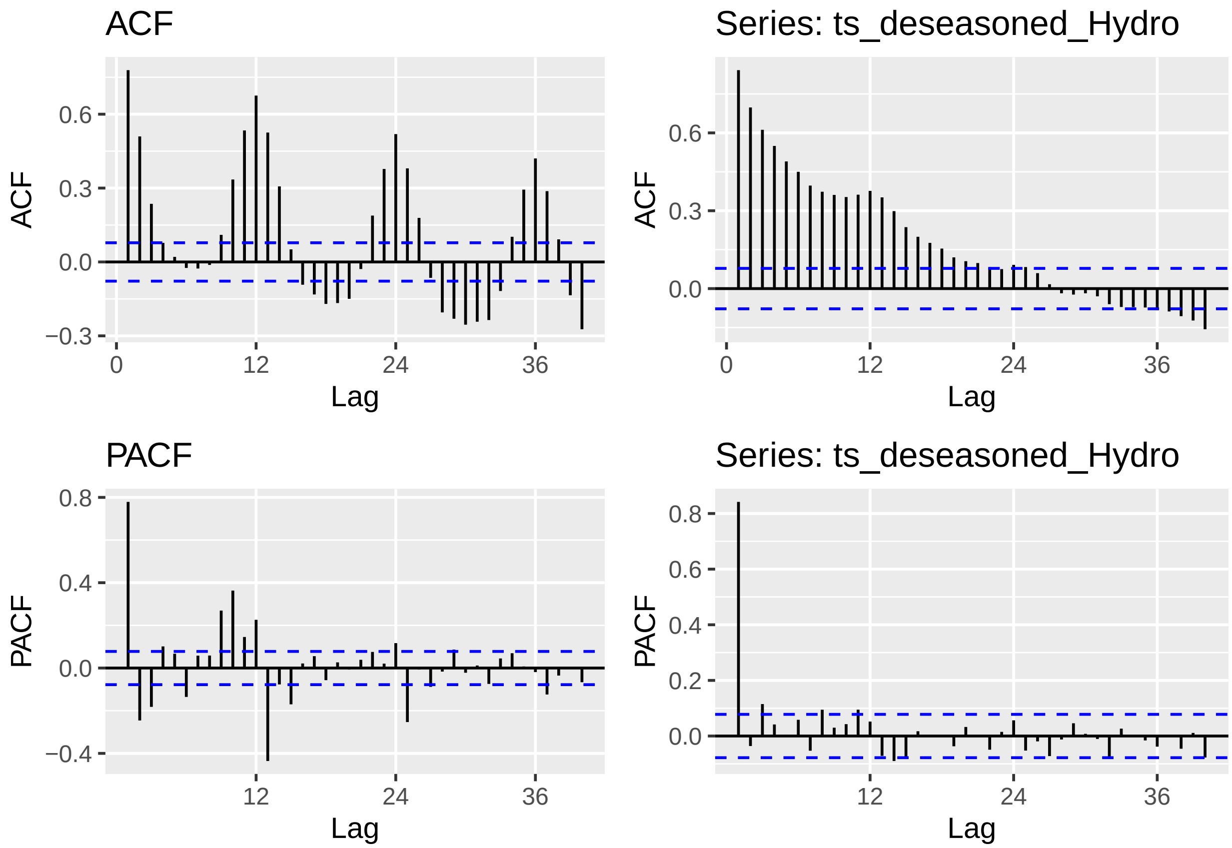

Series ts_deseasoned_renewables
Partial ACF
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
12
24
36
Lag




```
cowplot::plot_grid(Hydro2, autoplot(Acf(ts_deseasoned_Hydro, lag.max = 40)),
                   Hydro3, autoplot(Pacf(ts_deseasoned_Hydro, lag.max = 40)),
                   nrow = 2, ncol = 2)
```


Series ts_deseasoned_Hydro
ACF
0.0
0.2
0.4
0.6
0.8
0
12
24
36
Lag


Series ts_deseasoned_Hydro
Partial ACF
0.0
0.2
0.4
0.6
0.8
12
24
36
Lag



Answer: The total renewables time series did not change significantly, but we note that the PACF becomes insignificant much more quickly with deseasoned data than in the original time series. The hydroelectric ACF looks remarkably different, with a much lower sign of seasonality and a more gradual tapering of significant lags. The PACF looks different as well, indicating that the first lag is most important, with some debatably significant lags at 3, 8 and 11.