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

Assignment 3 - Due date 02/03/26

Maeve Gualtieri-Reed

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 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.\

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

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

```r
library(tseries)
library(Kendall)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(readxl)
library(lubridate)
```

```
##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
library(ggplot2)
```

```
#Importing data set
energy_data <- read_excel("/home/guest/TSA_Sp26/Data/Table_10.1_Renewable_Energy_Production_and_Consump
```

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

```
#Now let's extract the column names from row 11
read_col_names <- read_excel("/home/guest/TSA_Sp26/Data/Table_10.1_Renewable_Energy_Production_and_Consu
```

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

```
## * `` -> `...6`
## * `` -> `...7`
## * `` -> `...8`
## * `` -> `...9`
## * `` -> `...10`
## * `` -> `...11`
## * `` -> `...12`
## * `` -> `...13`
## * `` -> `...14`
```

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

#make date column a date data type
energy_data$Month <- as.Date(ymd(energy_data$Month))

#Create a data frame structure with these two time series only.
Renewable_Hydro <- energy_data %>%
  select('Month','Total Renewable Energy Production','Hydroelectric Power Consumption')
```

##Trend Component

Q1: TS, ACF, PACF

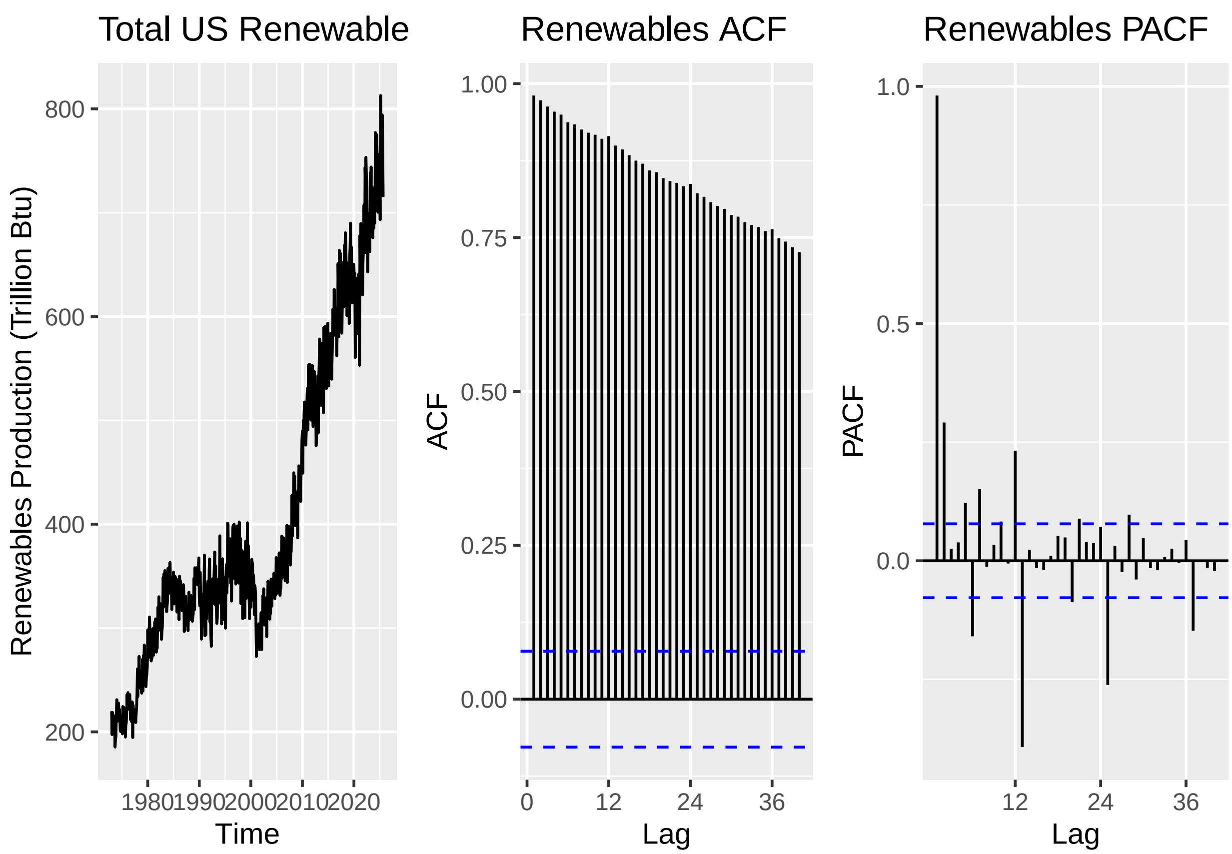
For each series (Total Renewable Production and Hydroelectric Consumption) create three plots arranged in a row (side-by-side): (1) time series plot, (2) ACF, (3) PACF. Use cowplot::plot_grid() to place them in a grid.

```
#time series plot

ts_Rewenable_Hydro <- ts(Renewable_Hydro[,2:3],
                     start=c(1973,1),
                     frequency=12)

renew_ts_plot <- autoplot(ts_Rewenable_Hydro[,1]) +
  xlab("Time") +
  ylab("Renewables Production (Trillion Btu)")+
  ggtitle("Total US Renewable Energy Production")

hydro_ts_plot <- autoplot(ts_Rewenable_Hydro[,2]) +
  xlab("Time") +
  ylab("Hydroelectric Consumption (Trillion Btu)")+
  ggtitle("US Hydroelectric Power Consumption")
```

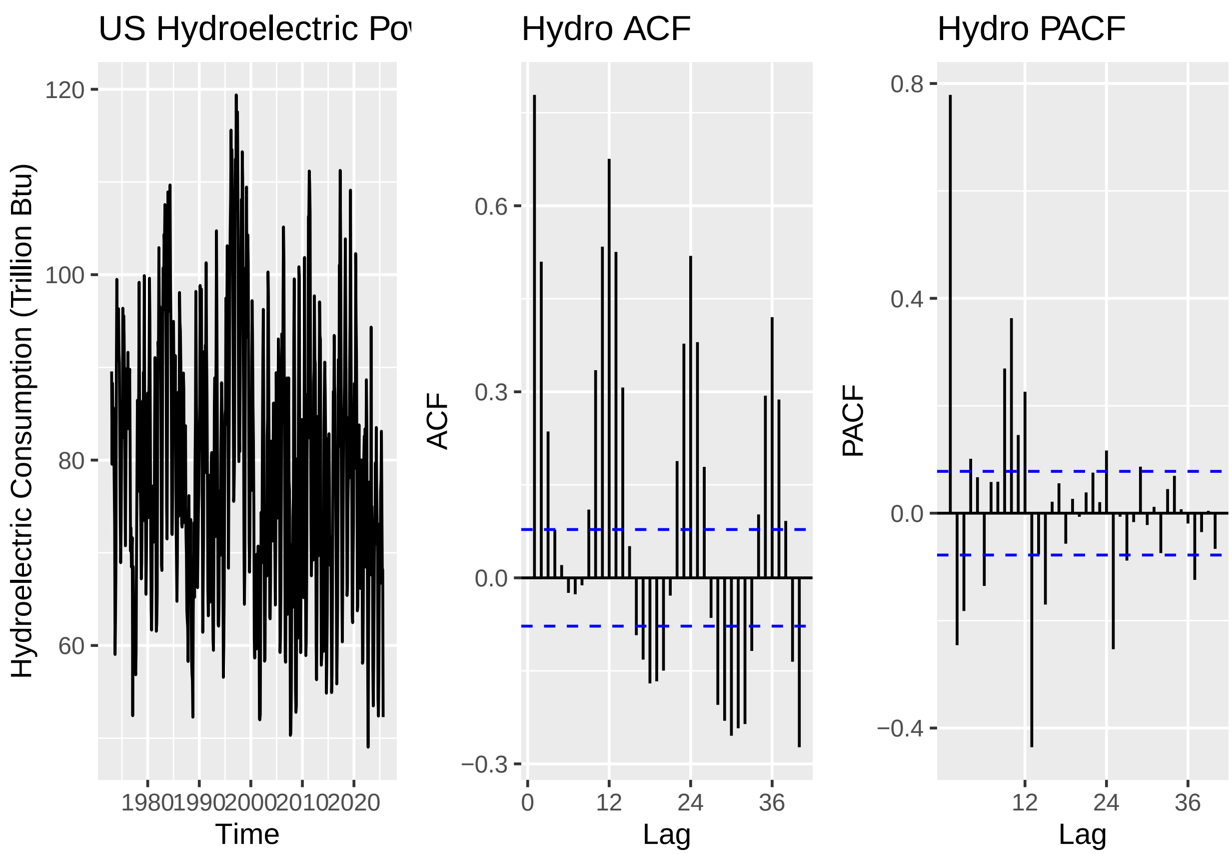
```
#ACF
#I use ggAcf and ggPacf so that I can plot with cowplot
renew_acf= ggAcf(ts_Rewenable_Hydro[,1], lag.max = 40)+
  ggtitle("Renewables ACF")

hydro_acf= ggAcf(ts_Rewenable_Hydro[,2], lag.max = 40)+
    ggtitle("Hydro ACF")
```

```r
renew_pacf= ggPacf(ts_Rewenable_Hydro[,1], lag.max = 40)+
    ggtitle("Renewables PACF")
hydro_pacf= ggPacf(ts_Rewenable_Hydro[,2], lag.max = 40)+
    ggtitle("Hydro PACF")
```

```r
#renewables
cowplot::plot_grid(renew_ts_plot,renew_acf,renew_pacf, ncol = 3)
```




```r
#hydro
cowplot::plot_grid(hydro_ts_plot,hydro_acf,hydro_pacf, ncol = 3)
```



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?

Renewable Energy Production seems to have a seasonal trend and an overall trend (increasing). There are regular intervals of peaks and troughs indicating the seasonal element. Overall the production is increasing from the start to the end of the series. While it is not visible in the ACF, the seasonal trend component is visible in the PACF as there are peaks around the annual mark.

Hydroelectric Power Consumption seems to have a clear seasonal trend. There are clear peaks and troughs in the timeseries plot, ACF, and PACF. There might also be a slight overall decreasing trend.

Q3: lm()

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.

```
t <- c(1:nrow(Renewable_Hydro))

renewable_lm <- lm(Renewable_Hydro$`Total Renewable Energy Production` ~ t)
summary(renewable_lm)
```

```
##
```

```
## Call:
## lm(formula = Renewable_Hydro$‘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
renew_int <- as.numeric(renewable_lm$coefficients[1])
renew_slope <- as.numeric(renewable_lm$coefficients[2])

hydro_lm <- lm(Renewable_Hydro$`Hydroelectric Power Consumption` ~ t)
summary(hydro_lm)
```

```
##
## Call:
## lm(formula = Renewable_Hydro$‘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
hydro_int <- as.numeric(hydro_lm$coefficients[1])
hydro_slope <- as.numeric(hydro_lm$coefficients[2])
```

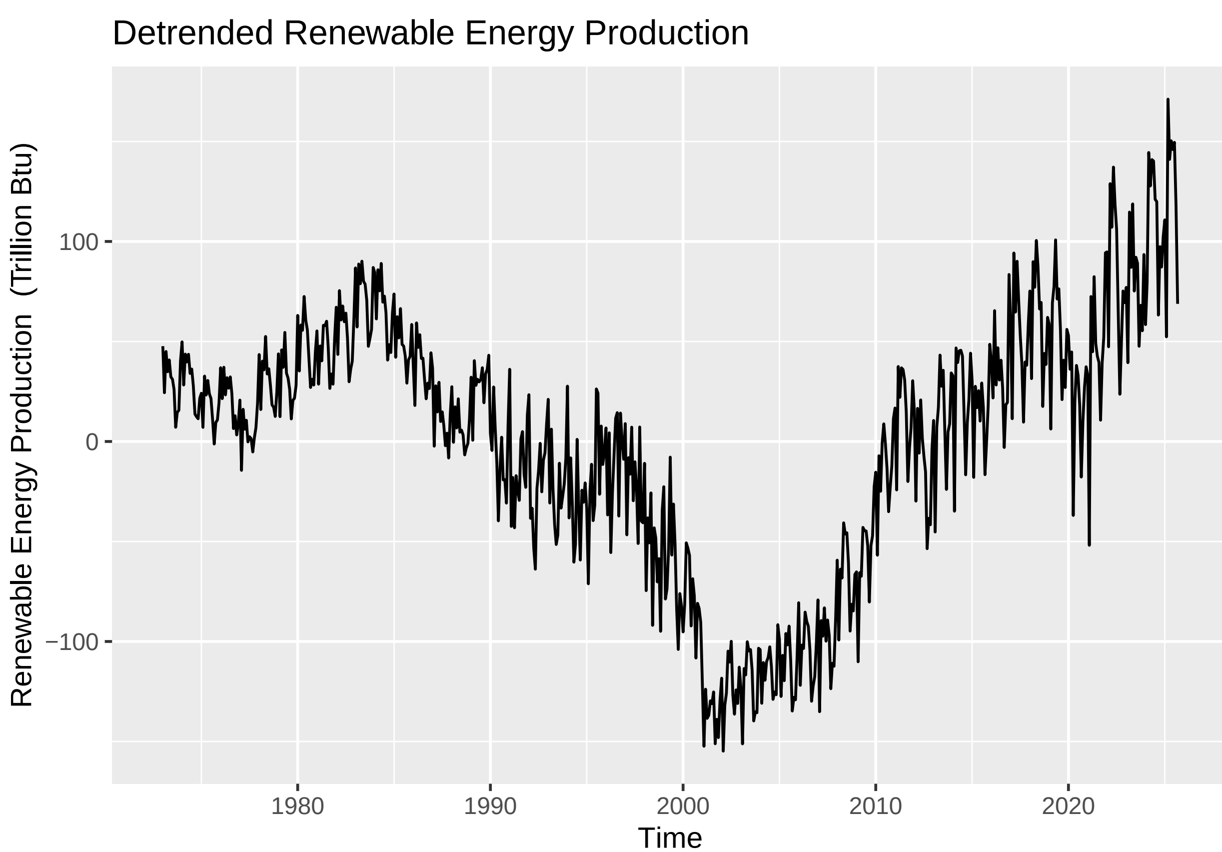
As expected, the renewable energy production has a positive slope (~0.75). The p value is significant, indicating that there is a significant increase in renewable production over time and that we can use the linear regression model to detrend the series.

The hydroelectric consumption linear regression also resulted in a significant p value for the slope value, though the slope value is negative and smaller (-.012). This indicates that there is a significant slight decrease in hydroelectric power consumption over time and that we can use the linear regression model to detrend the series.

Q4 detrend

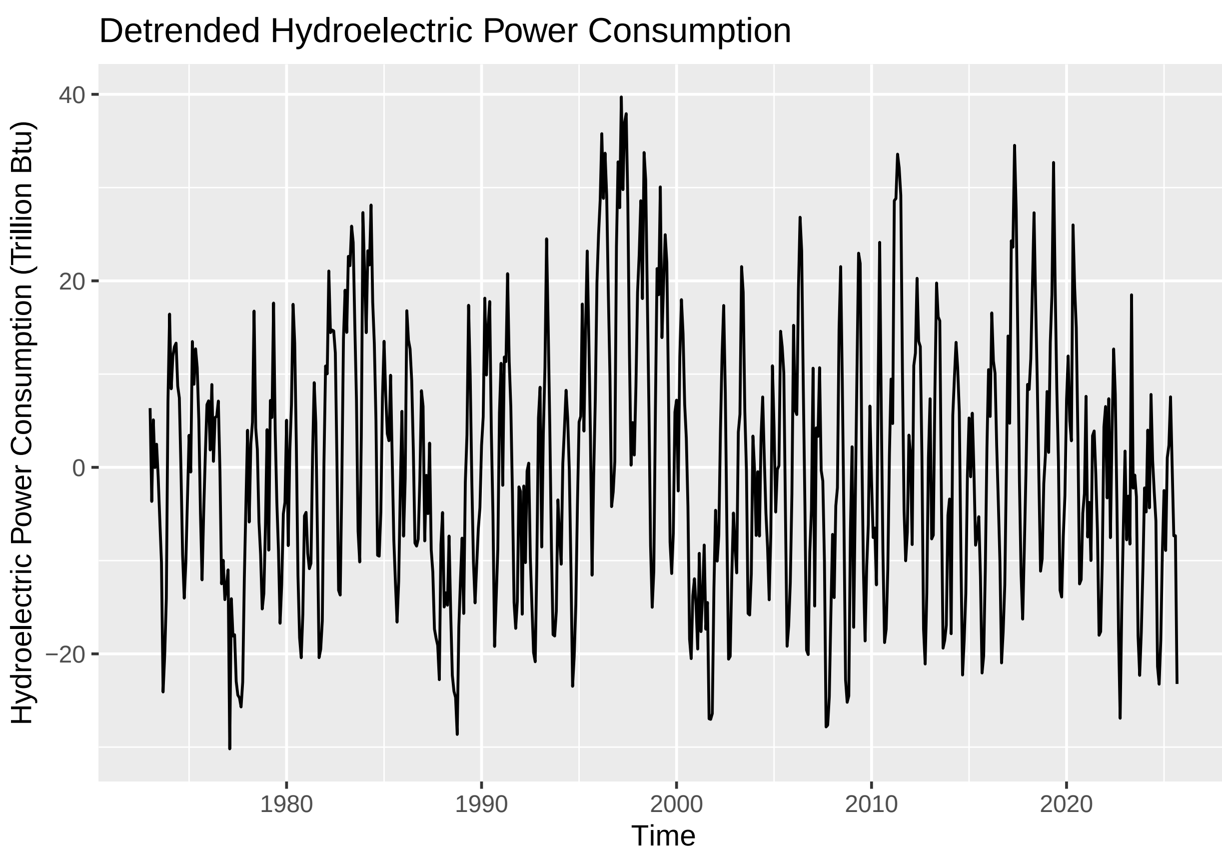
Use the regression coefficients to detrend each series (subtract fitted linear trend). Plot detrended series and compare with the original time series from Q1. Describe what changed.

```
#renewables
detrend_renew <- Renewable_Hydro[,2] - (renew_int+renew_slope*t)

ts_detrend_renew <- ts(detrend_renew,
                       start=c(1973,1),
                       frequency=12)

detrend_renew_plot<- autoplot(ts_detrend_renew) +
  xlab("Time") +
  ylab("Renewable Energy Production  (Trillion Btu)")+
  ggtitle("Detrended Renewable Energy Production")

detrend_renew_plot
```




```
#Hydro
detrend_hydro <- Renewable_Hydro[,3] - (hydro_int+hydro_slope*t)
```

```r
ts_detrend_hydro <- ts(detrend_hydro,
                       start=c(1973,1),
                       frequency=12)

detrend_hydro_plot<- autoplot(ts_detrend_hydro) +
  xlab("Time") +
  ylab("Hydroelectric Power Consumption (Trillion Btu)")+
  ggtitle("Detrended Hydroelectric Power Consumption")


detrend_hydro_plot
```



Both are now centered around 0 and do not show an overall increase or decrease. For renewables, there is a dip around 2000 in the detrended series rather than a steady increase over time seen in the original series. For hydro, There is much less variation in where the peaks and troughs sit and the slight decrease overtime in the original series is not as pronounced in the detrended series.

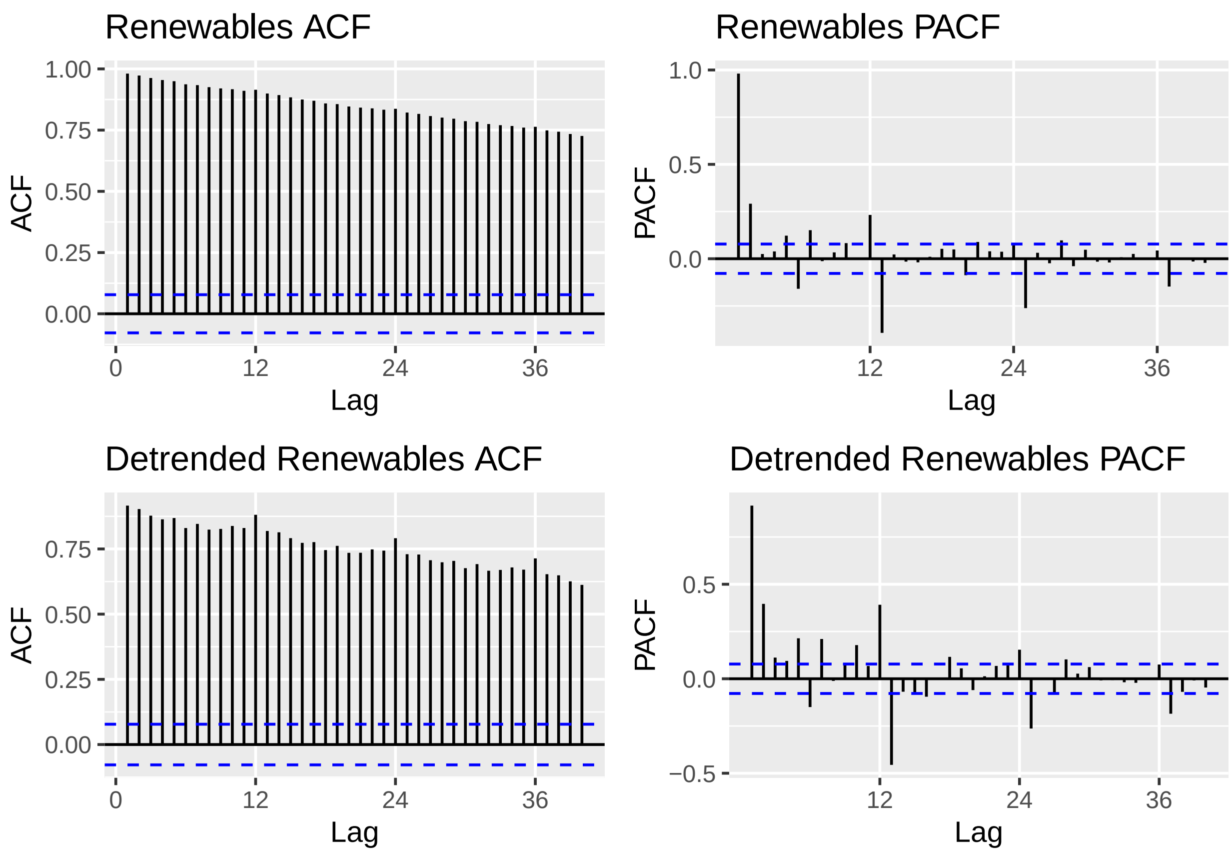
Q5 detrended ACF PACF

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?

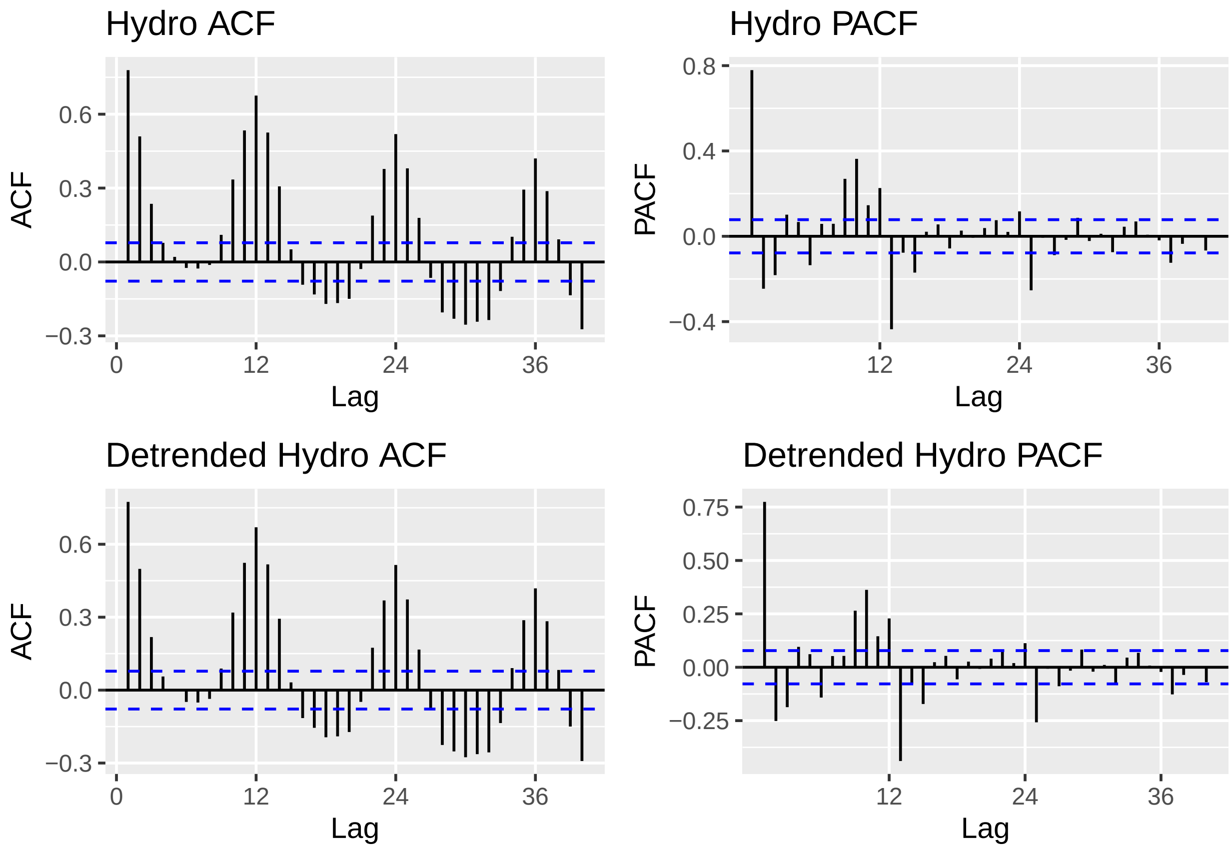
```r
detrend_renew_acf= ggAcf(ts_detrend_renew, lag.max = 40)+
  ggtitle("Detrended Renewables ACF")
```

```r
detrend_hydro_acf= ggAcf(ts_detrend_hydro, lag.max = 40)+
    ggtitle("Detrended Hydro ACF")

detrend_renew_pacf= ggPacf(ts_detrend_renew, lag.max = 40)+
  ggtitle("Detrended Renewables PACF")

detrend_hydro_pacf= ggPacf(ts_detrend_hydro, lag.max = 40)+
    ggtitle("Detrended Hydro PACF")
```

```r
cowplot::plot_grid(renew_acf, renew_pacf, detrend_renew_acf,detrend_renew_pacf,ncol = 2)
```




```r
cowplot::plot_grid(hydro_acf, hydro_pacf, detrend_hydro_acf,detrend_hydro_pacf,ncol = 2)
```



There is not a significant difference in the ACF or the PACF for renewables or between the detrended and raw data. This indicates that the trend wasn't hiding a significant seasonal component and that instead maybe there is not a significant seasonal component.

The hydro detrended vs original PACF and ACF are also fairly similar. We can see that there is a clear seasonal component in the detrended and raw ACF/PACF so it might be that the trend component wasn't strong enough to hide additional seasonal correlation. This makes sense as while there was a significant coefficient found in the linear regression, it was very very slight. ## 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 hydroelectric consumption timeseries definitely seems to have a seasonal trend. There are regular peaks and troughs equally spaced in the raw data. The ACF and PACF further support this as they both show strong correlation large lags. The renewable energy consumption is not as obvious. The raw data shows some regular peaks and troughs but the ACF shows all positive values with a very slow decline. However, just by looking at the ACF that includes the trend, it is possible that the seasonality is overshadowed by the strong overall trend computed in Q3. The PACF shows that there is likely some seasonality as some of the large lags show high positive and negative values.

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?

```
renew_dummies <- seasonaldummy(ts_Rewenable_Hydro[,1])
renew_seasonal_means <- lm(ts_Rewenable_Hydro[,1]~renew_dummies)
summary(renew_seasonal_means)
```

```
##
## Call:
## lm(formula = ts_Rewenable_Hydro[, 1] ~ renew_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 ***
## renew_dummiesJan    2.090     29.693   0.070    0.944
## renew_dummiesFeb  -34.524     29.693  -1.163    0.245
## renew_dummiesMar    5.956     29.693   0.201    0.841
## renew_dummiesApr   -6.900     29.693  -0.232    0.816
## renew_dummiesMay    8.162     29.693   0.275    0.784
## renew_dummiesJun   -2.231     29.693  -0.075    0.940
## renew_dummiesJul    3.864     29.693   0.130    0.897
## renew_dummiesAug   -3.978     29.693  -0.134    0.893
## renew_dummiesSep  -29.033     29.693  -0.978    0.329
## renew_dummiesOct  -19.937     29.834  -0.668    0.504
## renew_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
```

```
hydro_dummies <- seasonaldummy(ts_Rewenable_Hydro[,2])
hydro_seasonal_means <- lm(ts_Rewenable_Hydro[,2]~renew_dummies)
summary(hydro_seasonal_means)
```

```
##
## Call:
## lm(formula = ts_Rewenable_Hydro[, 2] ~ renew_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 ***
## renew_dummiesJan    4.951      2.021   2.449 0.014591 *
## renew_dummiesFeb   -2.415      2.021  -1.195 0.232608
## renew_dummiesMar    7.116      2.021   3.520 0.000463 ***
## renew_dummiesApr    5.614      2.021   2.777 0.005649 **
## renew_dummiesMay   14.080      2.021   6.965 8.38e-12 ***
## renew_dummiesJun   10.780      2.021   5.333 1.36e-07 ***
## renew_dummiesJul    4.003      2.021   1.980 0.048091 *
## renew_dummiesAug   -5.320      2.021  -2.632 0.008710 **
## renew_dummiesSep  -16.598      2.021  -8.211 1.28e-15 ***
## renew_dummiesOct  -16.329      2.031  -8.040 4.56e-15 ***
## renew_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
```

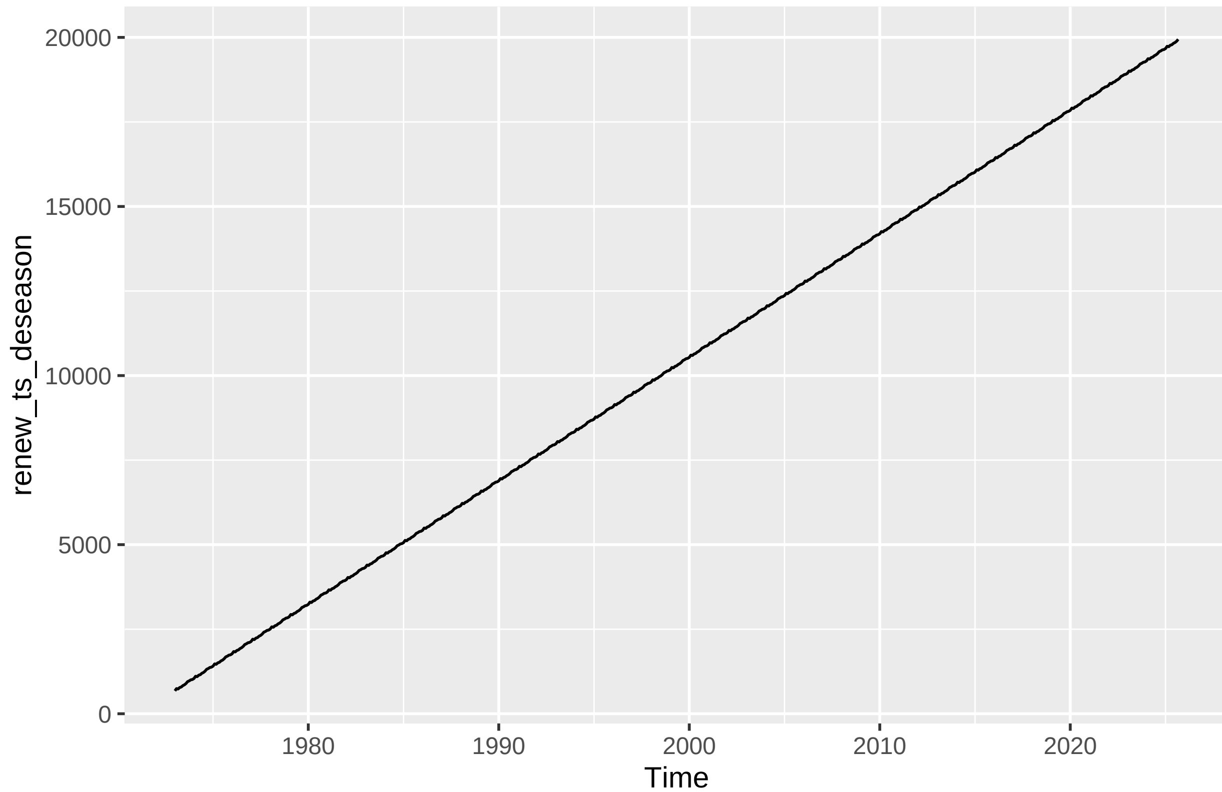
The seasonal means model for the hydro data shows an overall p value of less than 0.05 and all months' coefficients are < 0.05 except for February. Therefore, we can conclude that the series has a seasonal component. This matches what I inferred in Q6.

Renewable seasonal means model regression did not have a p value <0.05 and none of the coefficients were signficiant. We can not conclude that the series has a seasonal component. I thought that the renewable time series had a seasonal component.

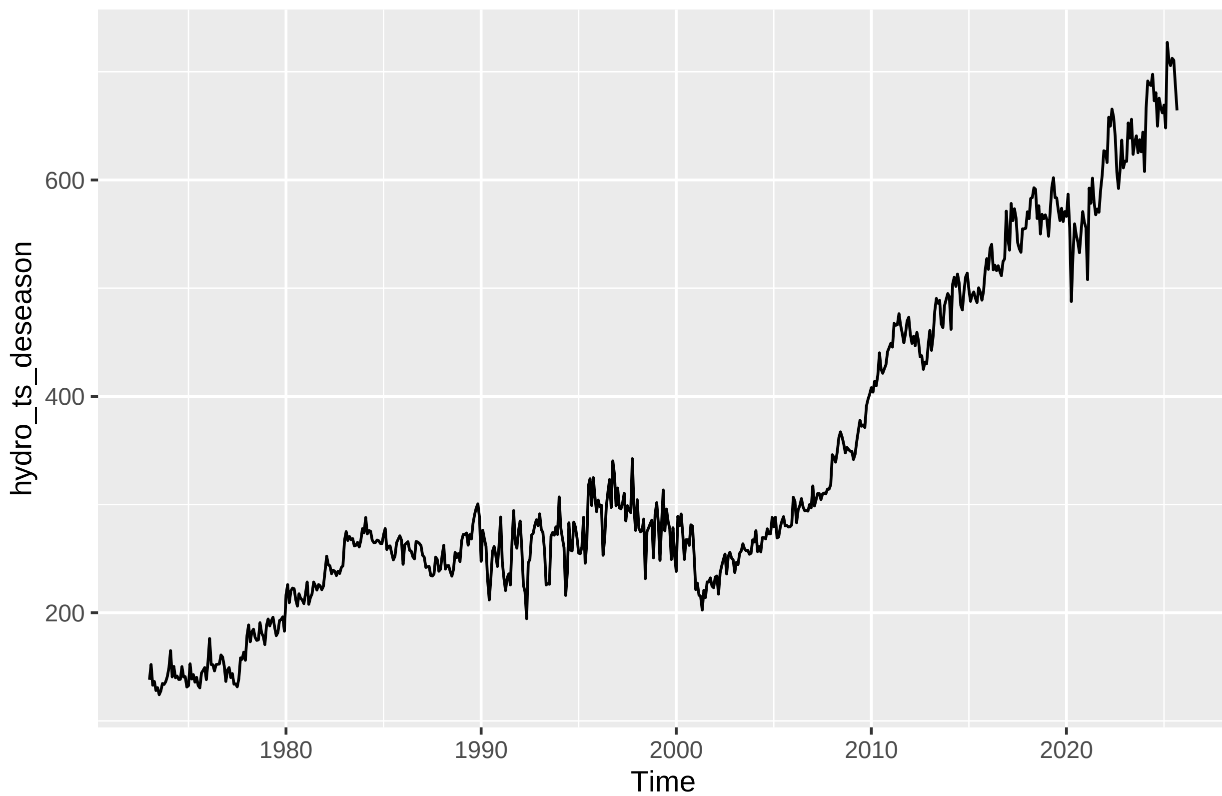
Q8

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
#store coeff
renew_int <- renew_seasonal_means$coefficients[1]
renew_coeff <-renew_seasonal_means$coefficients[2:12]

hydro_int <- hydro_seasonal_means$coefficients[1]
hydro_coeff <-hydro_seasonal_means$coefficients[2:12]

#renew seasonal
renew_seasonal <- array(0,nrow(Renewable_Hydro))

for(i in 1:nrow(renew_seasonal)){
  renew_seasonal[i] <- renew_int+renew_coeff%*%renew_dummies[i,]
}

#remove renew seasonal
renew_deseason <- Renewable_Hydro[1] - renew_seasonal

renew_ts_deseason <- ts(renew_deseason, start=c(1973,1),
                        frequency=12)

autoplot(renew_ts_deseason)
```




```r
#hydro seasonal
hydro_seasonal <- array(0,nrow(Renewable_Hydro))

for(i in 1:nrow(hydro_seasonal)){
  hydro_seasonal[i] <- hydro_int+renew_coeff%*%hydro_dummies[i,]
}

#remove hydro seasonal
hydro_deseason <- Renewable_Hydro[2] - hydro_seasonal

hydro_ts_deseason <- ts(hydro_deseason, start=c(1973,1),
                        frequency=12)

autoplot(hydro_ts_deseason)
```



The renewable deseasoned plots is almost a perfectly straight upward line. There is not much variation aside from the upward trend without the seasonal component.

The hydro deseasoned plot still shows some variability but what is most prevalent is the clear upward overall trend seen without the seasonal component. This is interesting because with the raw time series, the linear regression showed a significant negative slope.

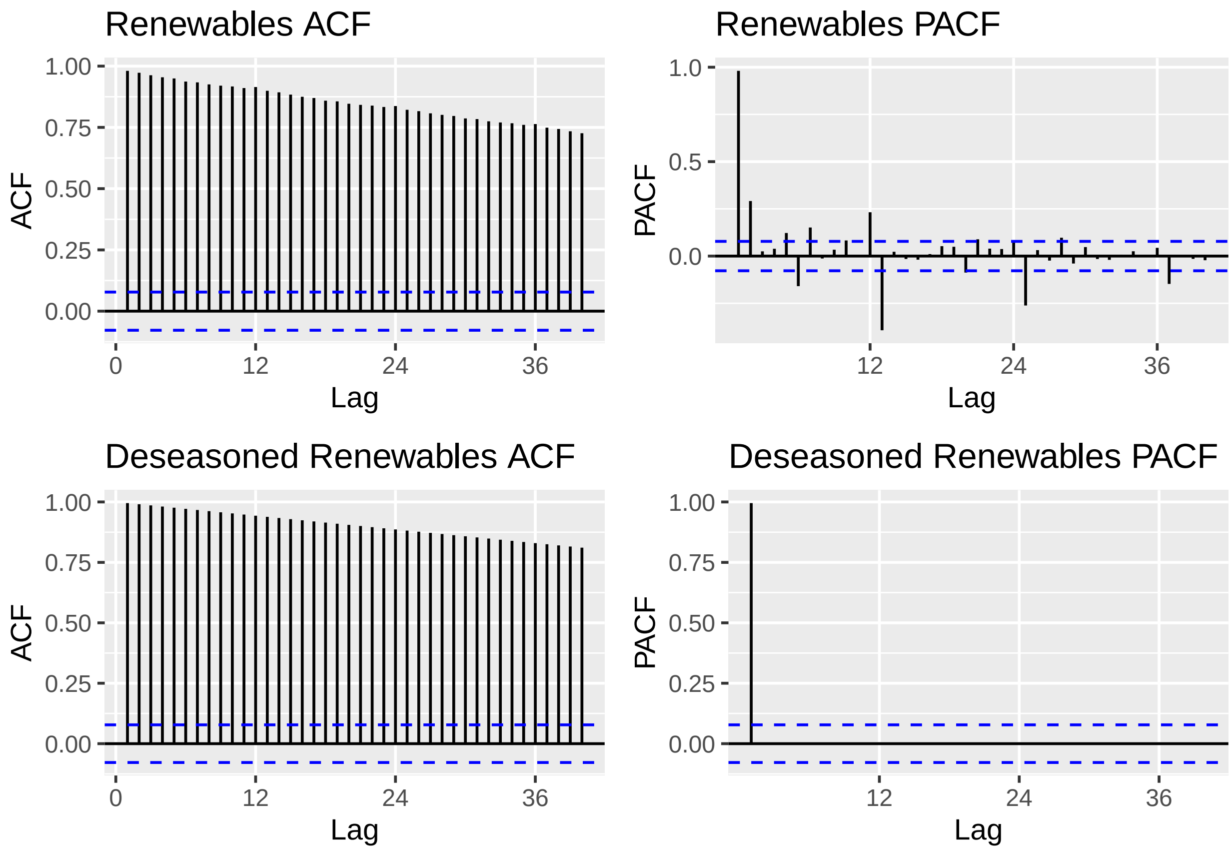
Q9

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?

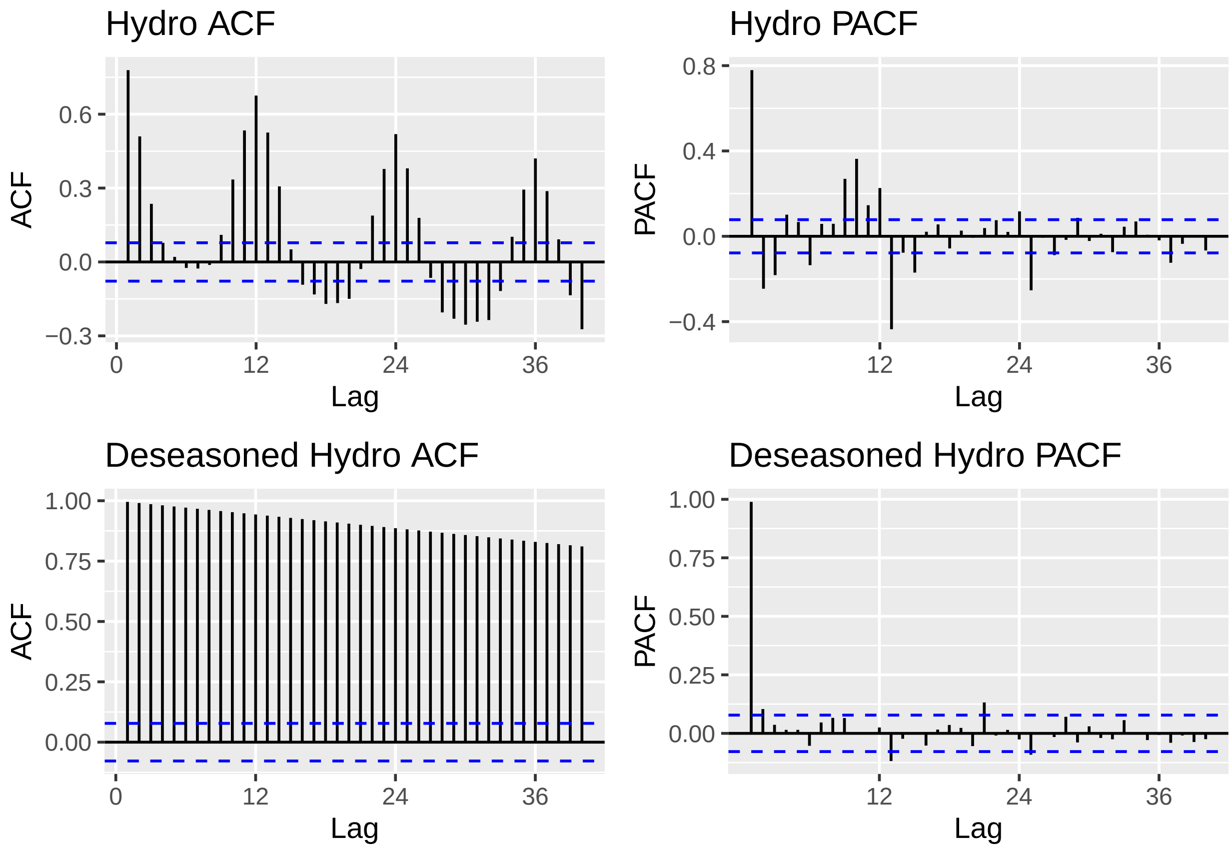
```r
#renew
deseas_renew_acf= ggAcf(renew_ts_deseason, lag.max = 40)+
  ggtitle("Deseasoned Renewables ACF")

deseas_renew_pacf= ggPacf(renew_ts_deseason, lag.max = 40)+
  ggtitle("Deseasoned Renewables PACF")
```

```r
#hydro
deseas_hydro_acf= ggAcf(renew_ts_deseason, lag.max = 40)+
  ggtitle("Deseasoned Hydro ACF")

deseas_hydro_pacf= ggPacf(hydro_ts_deseason, lag.max = 40)+
  ggtitle("Deseasoned Hydro PACF")
```

```r
#renewables
cowplot::plot_grid(renew_acf,renew_pacf, deseas_renew_acf,deseas_renew_pacf,ncol = 2)
```




```r
#hydro
cowplot::plot_grid(hydro_acf,hydro_pacf, deseas_hydro_acf,deseas_hydro_pacf,ncol = 2)
```



The renewables deseasoned ACF looks very similar to the raw ACF. All values are near 1 and there is a slow decline. The deseasoned PACF shows less variability, which makes sense since we have removed the seasonal component which likely caused the spikes in the raw PACF.

The hydro deseaseoned ACF looks very different. While in the raw ACF there was a clear sinusoidal shape, the deseasoned ACF has a slow decline and all positive values near 1. The deseasoned PACF still shows some variation, but not as much as the raw PACF, likely because the raw PACF was showing variability from the seasonal component.