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

Assignment 3 - Due date 02/04/25

Chloe Young

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(readxl)
library(openxlsx)
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
```

```r
library(lubridate)
```

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

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

```r
library(ggplot2)
library(cowplot)
```

```
##
## Attaching package: 'cowplot'

## The following object is masked from 'package:lubridate':
##
##     stamp
```

##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
#Importing data set

#Importing data set without change the original file using read.xlsx
energy_data1 <- read_excel(path="/home/guest/TSA_Sp25/Data/Table_10.1_Renewable_Energy_Production_and_C
```

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

```r
#Now let's extract the column names from row 11
read_col_names <- read_excel(path="/home/guest/TSA_Sp25/Data/Table_10.1_Renewable_Energy_Production_and_
```

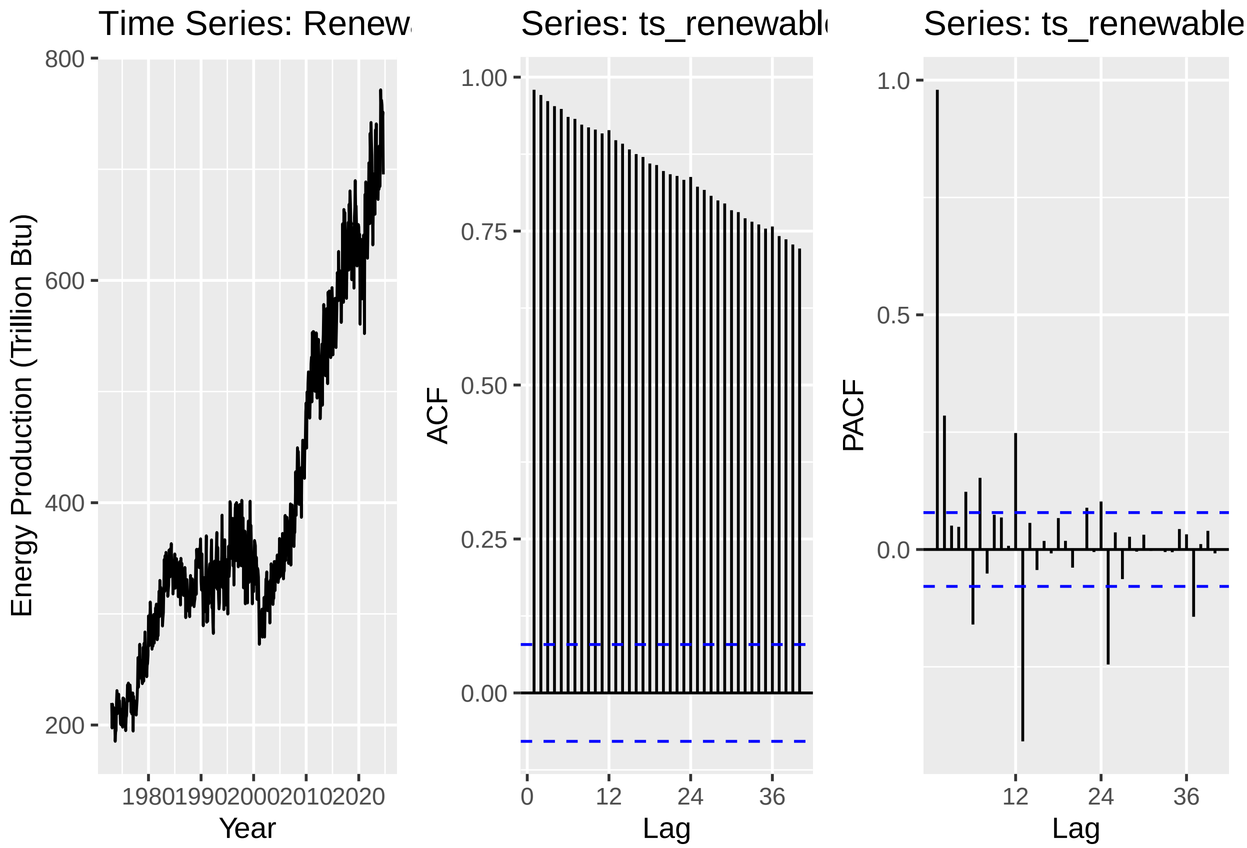
```
## New names:
## * `` -> `...1`
## * `` -> `...2`
## * `` -> `...3`
## * `` -> `...4`
## * `` -> `...5`
## * `` -> `...6`
## * `` -> `...7`
## * `` -> `...8`
## * `` -> `...9`
## * `` -> `...10`
## * `` -> `...11`
## * `` -> `...12`
## * `` -> `...13`
## * `` -> `...14`
```

```r
#Assign the column names to the data set
colnames(energy_data1) <- read_col_names

energy_data1 <- energy_data1[ ,5:6] #the space before the comma means you want all rows
                                  #and 4:6 means all columns from 4 to 6

#Visualize the first rows of the data set
head(energy_data1)
```

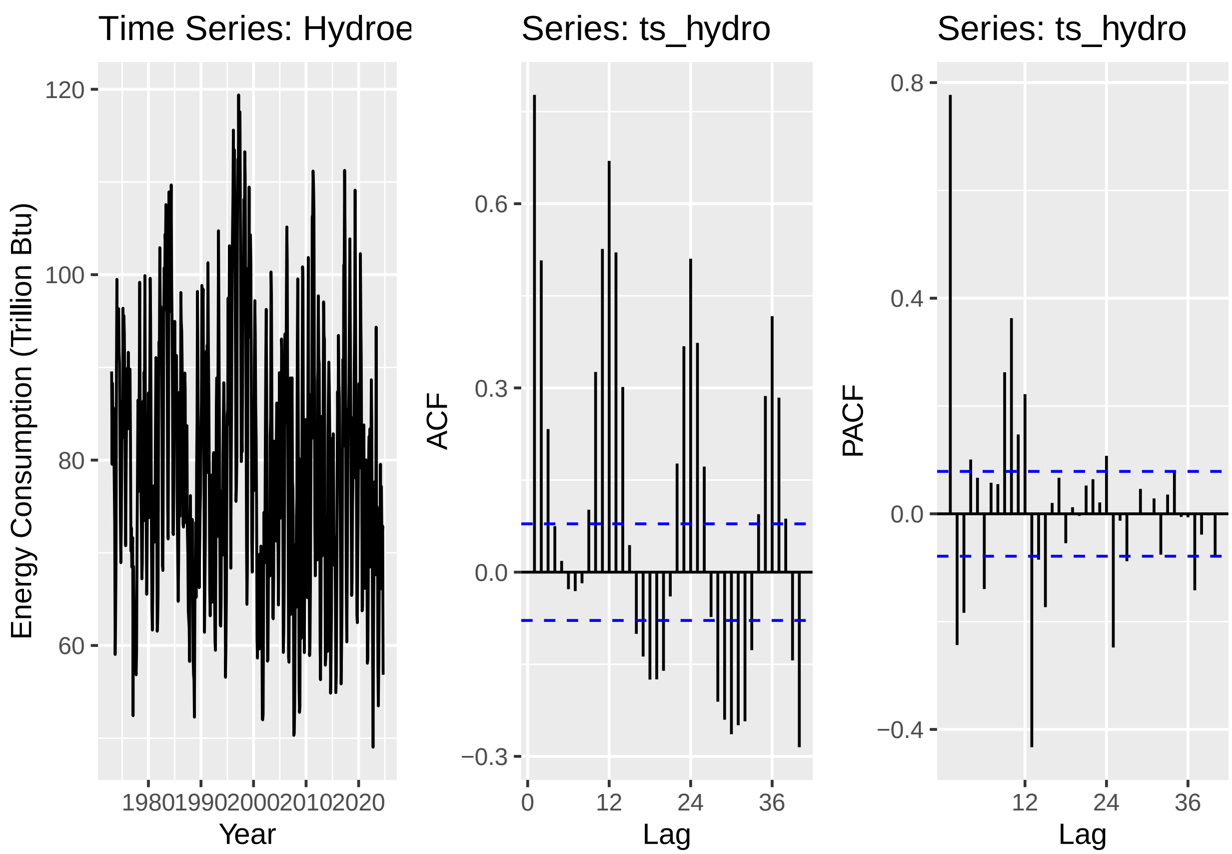
```
## # A tibble: 6 x 2
##   `Total Renewable Energy Production` `Hydroelectric Power Consumption`
##                                 <dbl>                             <dbl>
## 1                                220.                              89.6
## 2                                197.                              79.5
## 3                                219.                              88.3
## 4                                209.                              83.2
## 5                                216.                              85.6
## 6                                208.                              82.1
```

```r
nobs <- nrow(energy_data1)
```

```r
#transform into time series
ts_renewable <- ts(energy_data1$`Total Renewable Energy Production`, start=c(1973,1),frequency=12)
ts_hydro <- ts(energy_data1$`Hydroelectric Power Consumption`, start=c(1973,1),frequency=12)

#create time series plots
ts_r_plot <- autoplot(ts_renewable) +
  ggtitle("Time Series: Renewable Production") +
  xlab("Year") +
  ylab("Energy Production (Trillion Btu)") +
  labs(color="Reservoir")

ts_h_plot <- autoplot(ts_hydro) +
  ggtitle("Time Series: Hydroelectric Consumption") +
  xlab("Year") +
  ylab("Energy Consumption (Trillion Btu)") +
  labs(color="Reservoir")

#plot ACF
Renewable_acf=ggAcf(ts_renewable,lag.max=40, type="correlation", plot=TRUE)

Hydroelectric_acf=ggAcf(ts_hydro,lag.max=40, type="correlation", plot=TRUE)

#plot PACF
Renewable_pacf=ggPacf(ts_renewable,lag.max=40, plot=TRUE)

Hydroelectric_pacf=ggPacf(ts_hydro,lag.max=40, plot=TRUE)

#form into grid
plot_grid(ts_r_plot, Renewable_acf, Renewable_pacf, ncol = 3)
```




```r
plot_grid(ts_h_plot, Hydroelectric_acf, Hydroelectric_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?

Total Renewable Production appears to have an upward trend as energy production from renewables increases over time. There are slight dips in the data, such as the 2008 financial crisis and the 2020 COVID pandemic, but overall, production increases significantly.

Hydroelectric Power Consumption, on the other hand, doesn't appear to have a strong upward or downward trend, which is consistent with what we know about hydroelectric power since generation capacity hasn't increased substantially over the years.

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.

```
#create vector t
t <- c(1:nobs)

#combine t and original into one data frame
data_r <- data.frame("t"=t,"ts_data_r"=ts_renewable)
data_h <- data.frame("t"=t,"ts_data_h"=ts_hydro)
```

```r
#Fit a linear trend to TS, lm function needs a data frame object
linear_model_r=lm(ts_data_r~t,data_r)
linear_model_h=lm(ts_data_h~t,data_h)

summary(linear_model_r)
```

```
##
## Call:
## lm(formula = ts_data_r ~ t, data = data_r)
##
## 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_model_h)
```

```
##
## Call:
## lm(formula = ts_data_h ~ t, data = data_h)
##
## 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
```

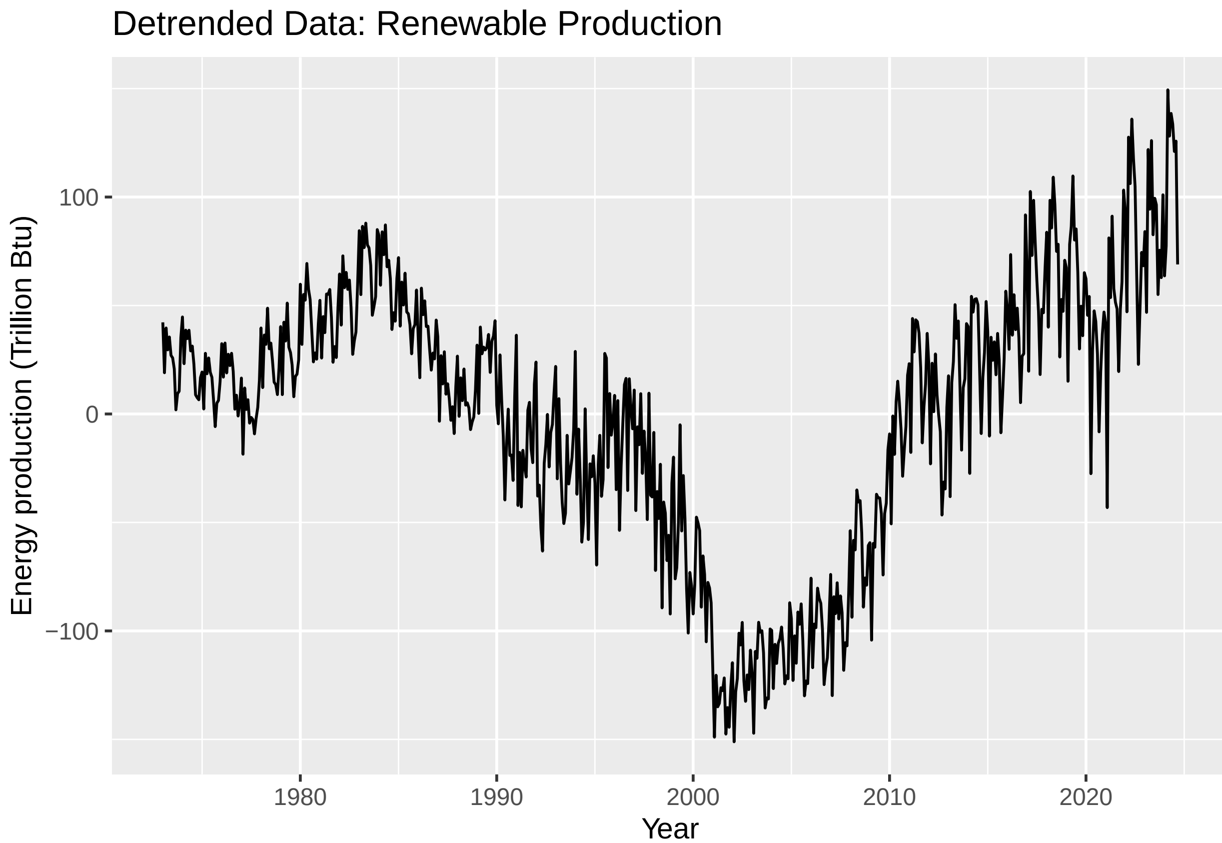
For renewable energy production, the p-value is 2.2e-16, which is less than 0.05, suggesting that there is statistical significance in the trend. Therefore, the null can be rejected. The slope is 0.72 which is positive and indicates that renewable power production is increasing over time, which is consistent with what we know about renewable energy.

For hydroelectric power consumption, the p-value is 0.00048, which is less than 0.05, suggesting that there is statistical significance in the trend, so the null can be rejected. The slope is -0.011, indicating a slight decreasing trend in hydroelectric power consumptio. However, this negative trend is very small.

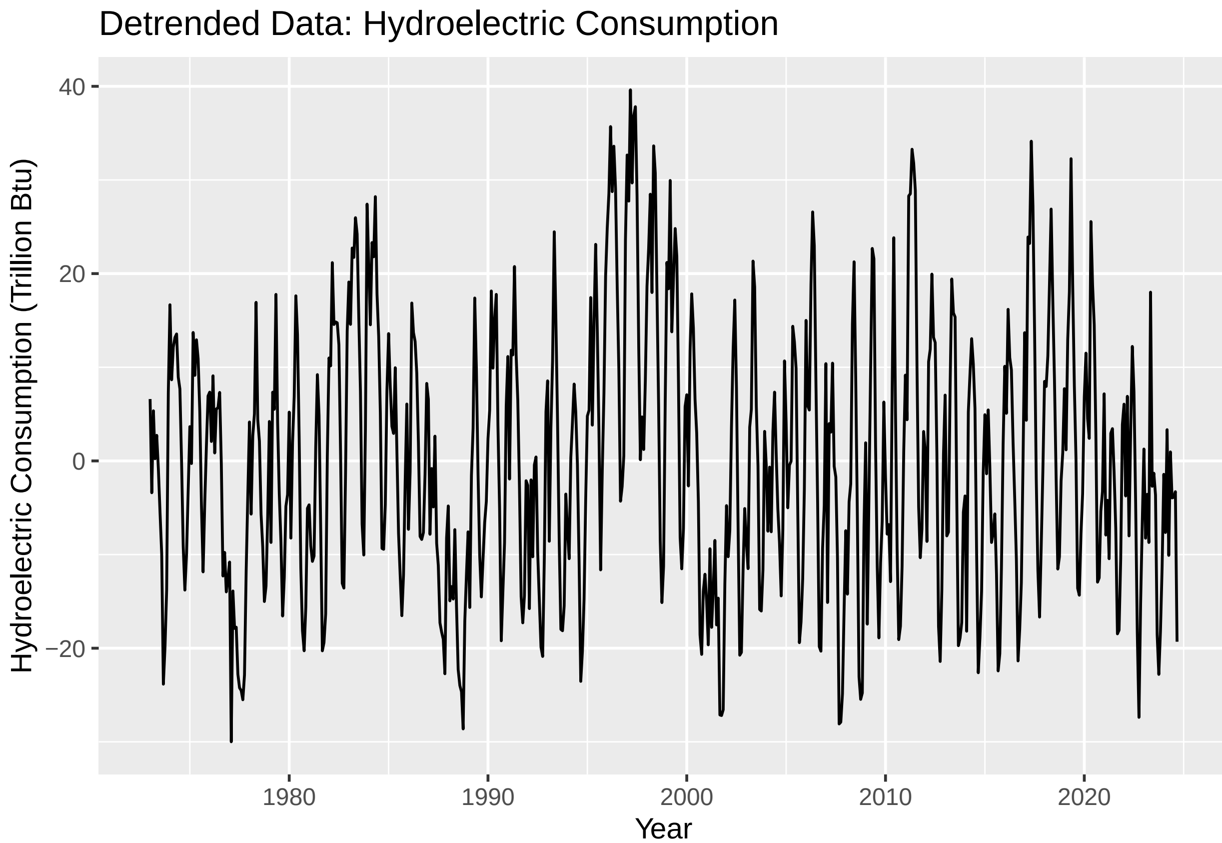
Q4

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?

```
beta0_r <- as.numeric(linear_model_r$coefficients[1])
beta1_r <- as.numeric(linear_model_r$coefficients[2])

beta0_h <- as.numeric(linear_model_h$coefficients[1])
beta1_h <- as.numeric(linear_model_h$coefficients[2])

#Detrend Renewable
linear_trend_r <- beta0_r + beta1_r * t
ts_linear_r <- ts(linear_trend_r,start=c(1973,1),frequency=12)

detrend_data_r <- ts_renewable - linear_trend_r
ts_r_detrend <- ts(detrend_data_r, start=c(1973,1),frequency = 12)

#Detrend Hydro
linear_trend_h <- beta0_h + beta1_h * t
ts_linear_h <- ts(linear_trend_h,start=c(1973,1),frequency=12)

detrend_data_h <- ts_hydro - linear_trend_h
ts_h_detrend <- ts(detrend_data_h, start = c(1973,1),frequency = 12)

#Plot
detrend_r_plot <- autoplot(ts_r_detrend) +
  ggtitle("Detrended Data: Renewable Production")+
  xlab("Year")+
  ylab("Energy production (Trillion Btu)")
detrend_r_plot
```




```
detrend_h_plot <- autoplot(ts_h_detrend) +
  ggtitle("Detrended Data: Hydroelectric Consumption")+
  xlab("Year")+
  ylab("Hydroelectric Consumption (Trillion Btu)")
detrend_h_plot
```



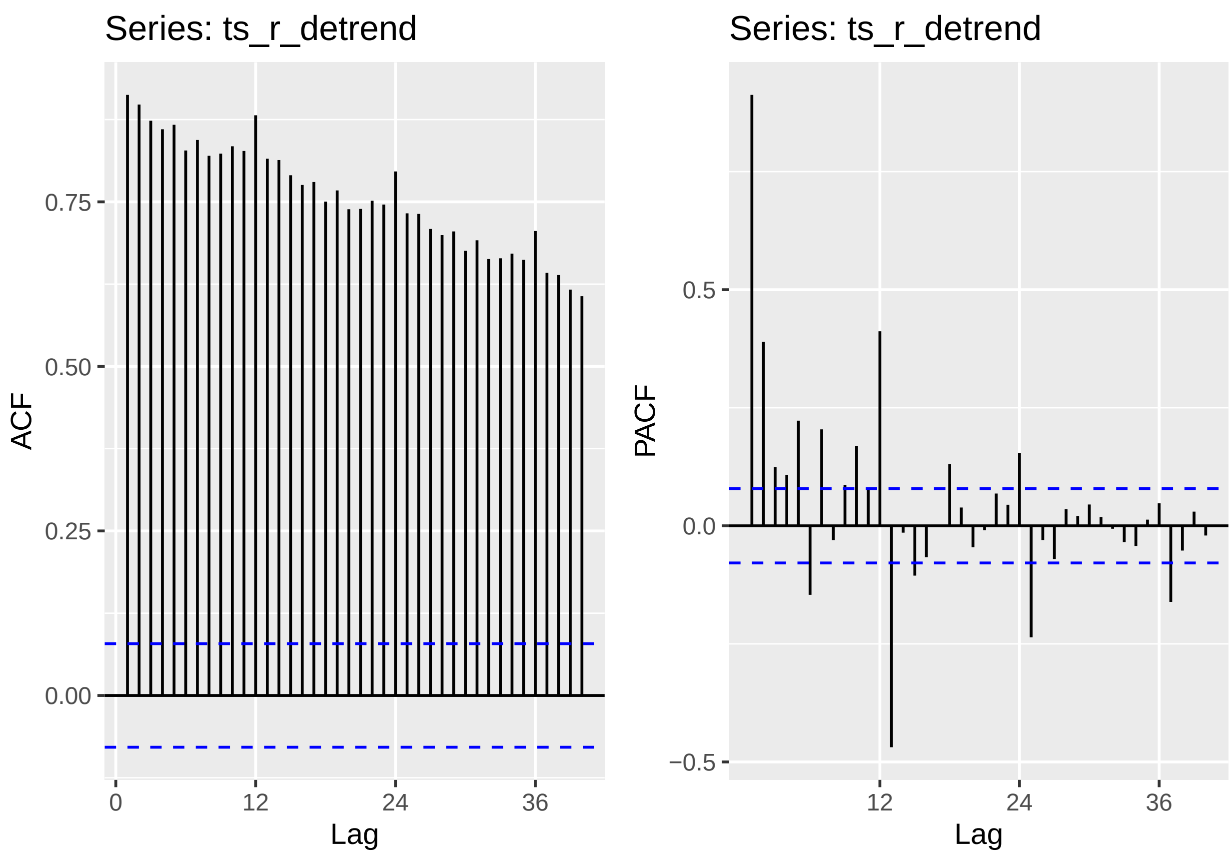
For renewable energy production, Q1 shows an increasing trend while Q4 shows less of this trend and instead stays closer to 0 with upwards and downwards fluctuations. This is likely because Q4 detrended the series, and without the trend, this long term increasing trend doesn't exist.

For hydroelectric power consumption, the data looks relatively similar but fluctuates around 0. There is no slight downward trend that was seen in Q1 since the trend was removed, but because the trend was so slight, it still looks relatively similar, although the fluctuations are no longer explained by the trend and are instead explained by seasonality or random variation.

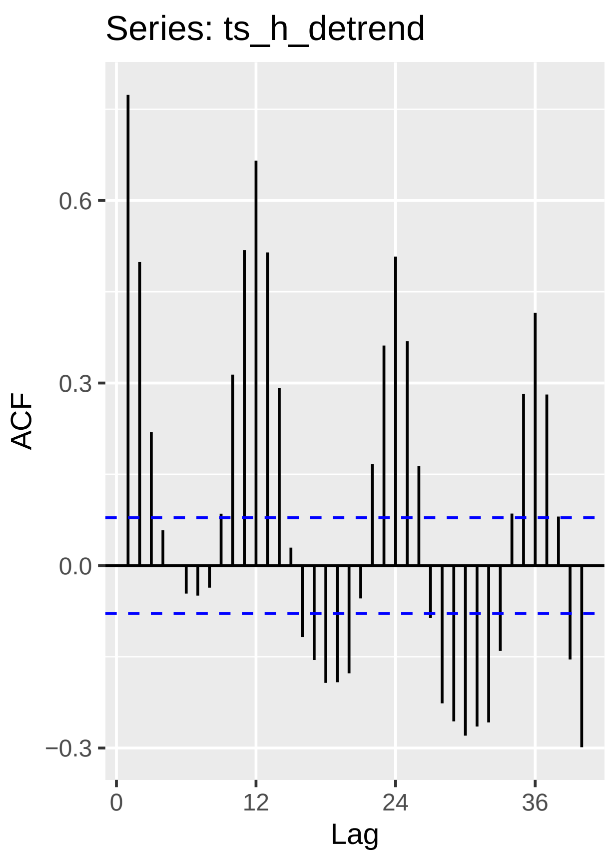
Q5

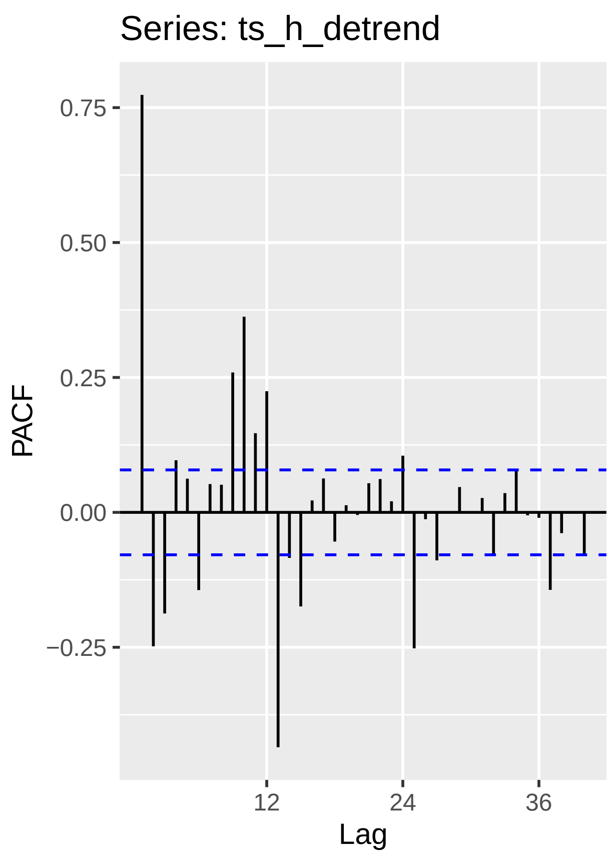
Plot ACF and PACF for the detrended series and compare with the plots from Q1. You may use plot_grid() again to get them side by side, but not mandatory. Did the plots change? How?

```
#plot ACF
DetrendRenewable_acf=ggAcf(ts_r_detrend, lag.max=40, type="correlation", plot=TRUE)

DetrendHydroelectric_acf=ggAcf(ts_h_detrend,lag.max=40, type="correlation", plot=TRUE)

#plot PACF
DetrendRenewable_pacf=ggPacf(ts_r_detrend,lag.max=40, plot=TRUE)

DetrendHydroelectric_pacf=ggPacf(ts_h_detrend,lag.max=40, plot=TRUE)

#form into grid
plot_grid(DetrendRenewable_acf, DetrendRenewable_pacf, ncol = 2)
```




```r
plot_grid(DetrendHydroelectric_acf, DetrendHydroelectric_pacf, ncol = 2)
```





For renewables, the ACF plot in Q1 has a more rapid decline than the plot in Q5. This is because the long term trend is removed so past values only influence future values over a short period. The PACF plot in Q1 shows a strong correlation at lag 1 and then on lags 13, 25, and 37, showing a seasonal trend. The PACF plot in Q5 is relatively similar.

For hydro, the ACF plots don't appear to have change much between questions 1 and 5, and this could be because the trend isn't strongly dependent on time so even after being detrended, the data looks similar. The ACF plot does indicate some cyclical variation in both. The PACF plots also look similar between both questions but show seasonal spikes at 13, 25, and 37 lags, indicating some potential seasonality.

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 Renewable Production, the time series doesn't indicate a seasonal trend. The ACF plot shows that there is autocorrelation and the gradual decline which indicates long term dependency in the data. However, the PACF shows spikes at 13, 25, and 37 months, indicating that there may be some sort of seasonal trend.

For Hydroelectric Consumption, The ACF and PACF plots indicate seasonal variation in the data as the lines spike every 12 months. This means that hydroelectric consumption is seasonal, which makes sense given that water availability from precipitation varies in different seasons.

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?

```
#Seasonal means model for Renewables
dummiesR <- seasonaldummy(ts_renewable)

seas_means_modelR <- lm(ts_renewable ~ dummiesR)
summary(seas_means_modelR)
```

```
##
## Call:
## lm(formula = ts_renewable ~ dummiesR)
##
## 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 ***
## dummiesRJan    1.973     28.468   0.069    0.945
## dummiesRFeb  -34.348     28.468  -1.207    0.228
## dummiesRMar    4.721     28.468   0.166    0.868
## dummiesRApr   -7.896     28.468  -0.277    0.782
## dummiesRMay    7.199     28.468   0.253    0.800
## dummiesRJun   -3.394     28.468  -0.119    0.905
## dummiesRJul    2.850     28.468   0.100    0.920
## dummiesRAug   -4.430     28.468  -0.156    0.876
## dummiesRSep  -29.037     28.468  -1.020    0.308
## dummiesROct  -20.205     28.606  -0.706    0.480
## dummiesRNov  -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
```

```
#Seasonal means model for Hydroelectric
dummiesH <- seasonaldummy(ts_hydro)

seas_means_modelH <- lm(ts_hydro ~ dummiesH)
summary(seas_means_modelH)
```

```
##
## Call:
## lm(formula = ts_hydro ~ dummiesH)
##
## 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 ***
## dummiesHJan    4.941      2.043   2.418 0.015880 *
## dummiesHFeb   -2.513      2.043  -1.230 0.219219
## dummiesHMar    7.053      2.043   3.452 0.000595 ***
## dummiesHApr    5.399      2.043   2.642 0.008441 **
## dummiesHMay   13.907      2.043   6.807 2.40e-11 ***
## dummiesHJun   10.729      2.043   5.251 2.09e-07 ***
## dummiesHJul    4.033      2.043   1.974 0.048845 *
## dummiesHAug   -5.399      2.043  -2.643 0.008436 **
## dummiesHSep  -16.596      2.043  -8.123 2.54e-15 ***
## dummiesHOct  -16.370      2.053  -7.973 7.66e-15 ***
## dummiesHNov  -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 renewable production, the p-value of 0.91 is greater than 0.05 indicating that there isn't statistical significance, so seasonality doesn't explain the data. The R^2 value is also very low at 0.0086, which indicates that seasonality explains less than 1% of the variability. Therefore, renewable production doesn't have a seasonal trend.

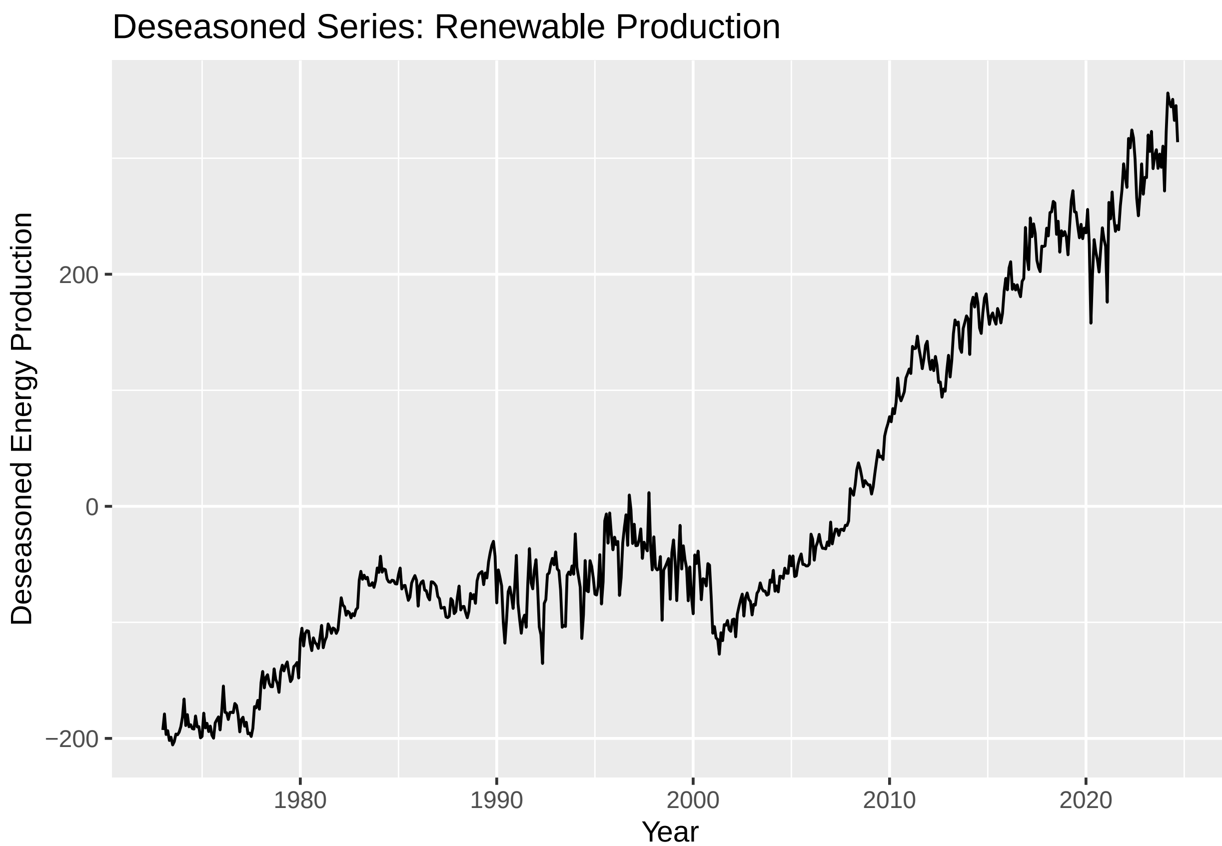
For hydroelectric consumption, the p-value is less than 0.05 so the null can be rejected as the data is statistically significant, so there is seasonality. The R^2 value is 0.46 indicating that seasonality explains 46% of the variability. Therefore, there is seasonality as hydroelectric consumption fluctuates throughout the year alongside environmental changes, as discussed in previous questions.

Q8

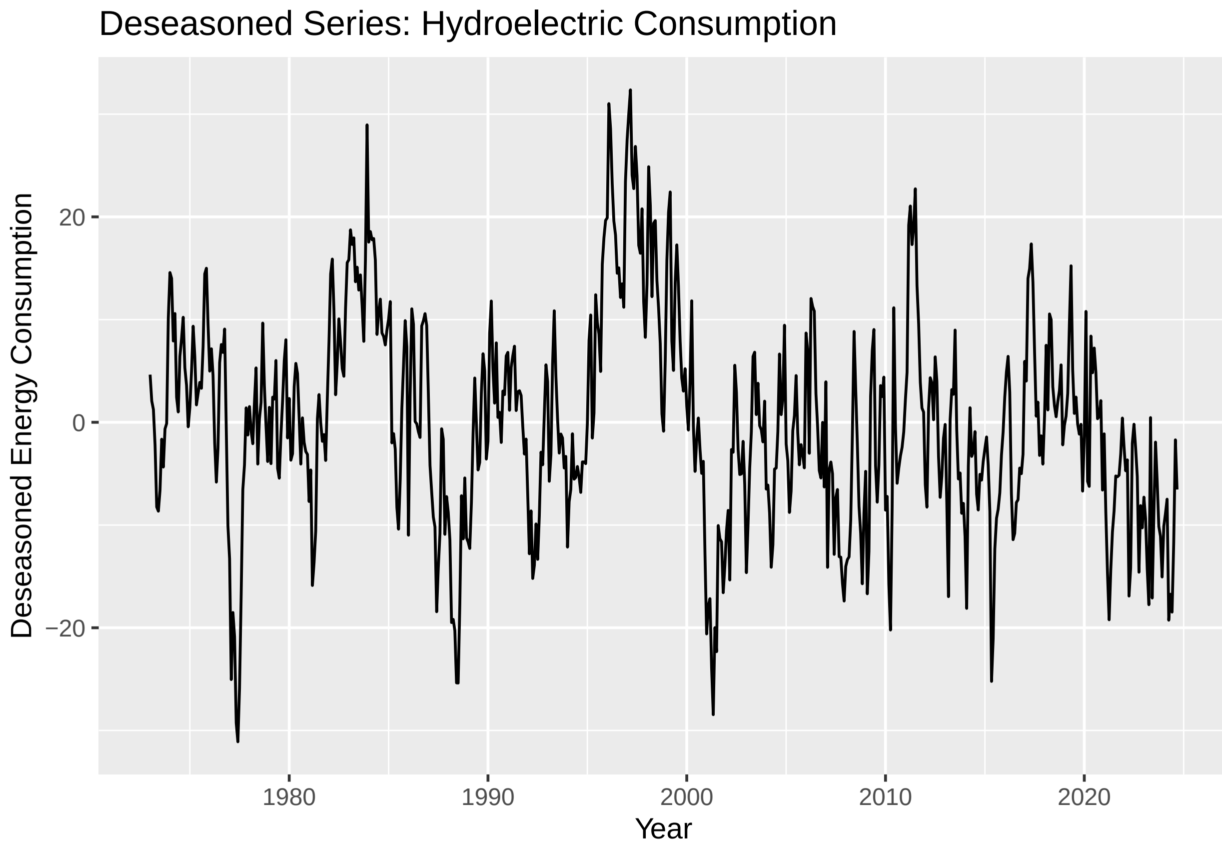
Use the regression coefficients from Q7 to deseason the series. Plot the deseason series and compare with the plots from part Q1. Did anything change?

```r
# Extract fitted seasonal component from the regression models
seasonal_component_R <- predict(seas_means_modelR)
seasonal_component_H <- predict(seas_means_modelH)

# Deseason
deseasoned_renewable <- ts_renewable - seasonal_component_R
deseasoned_hydro <- ts_hydro - seasonal_component_H

# Convert to TS
ts_deseasoned_renewable <- ts(deseasoned_renewable, start=c(1973,1), frequency=12)
ts_deseasoned_hydro <- ts(deseasoned_hydro, start=c(1973,1), frequency=12)

# Plots
deseasoned_r_plot <- autoplot(ts_deseasoned_renewable) +
  ggtitle("Deseasoned Series: Renewable Production") +
  xlab("Year") +
```

```r
  ylab("Deseasoned Energy Production")
deseasoned_r_plot
```




```r
deseasoned_h_plot <- autoplot(ts_deseasoned_hydro) +
  ggtitle("Deseasoned Series: Hydroelectric Consumption") +
  xlab("Year") +
  ylab("Deseasoned Energy Consumption")
deseasoned_h_plot
```



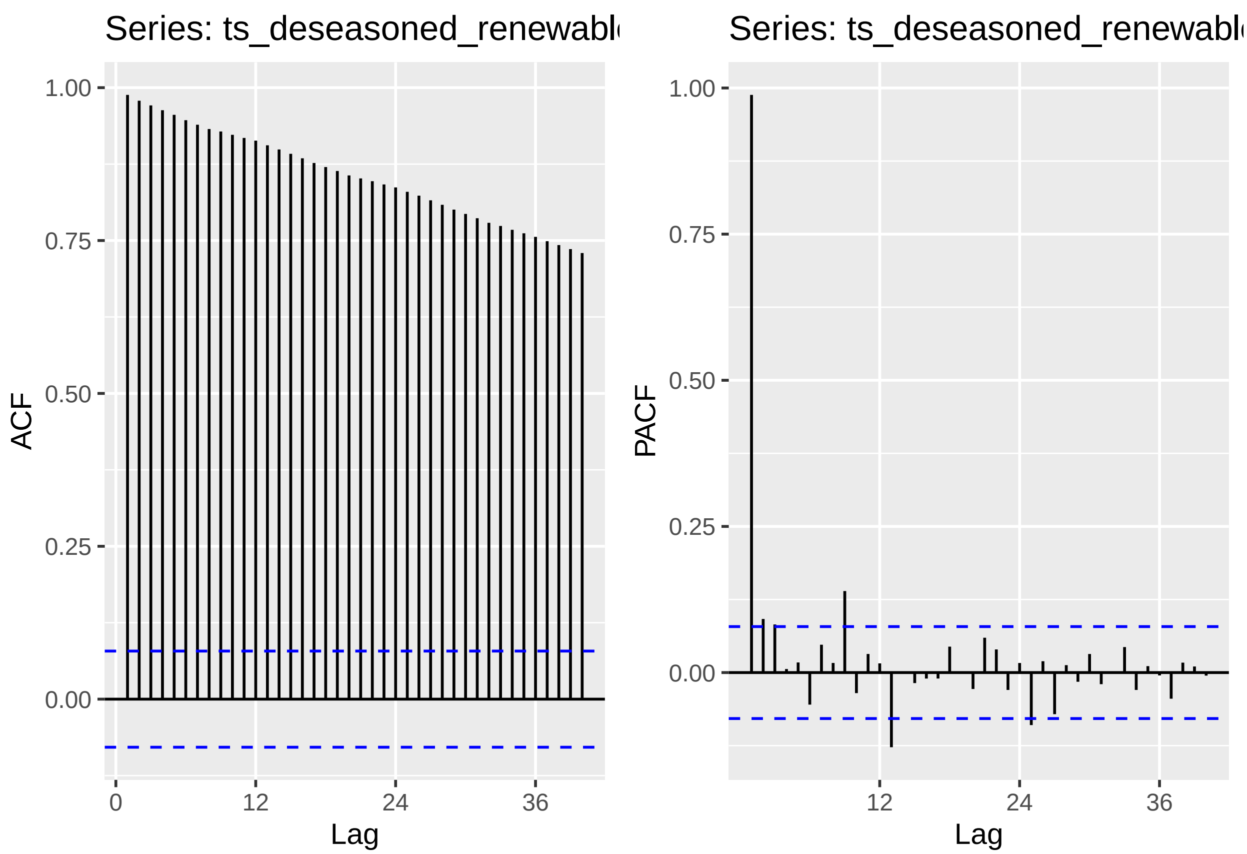
For renewable production, the deseasoned plot shows an upward trend, which is similar to Q1, indicating that the trend remains even when seasonality is removed, so the data isn't seasonal.

For hydroelectric consumption, the scale of the y axis is a lot smaller so the seasonal peaks have been reduced considerably since Q1, indicating that there is seasonality that has been removed from the data.

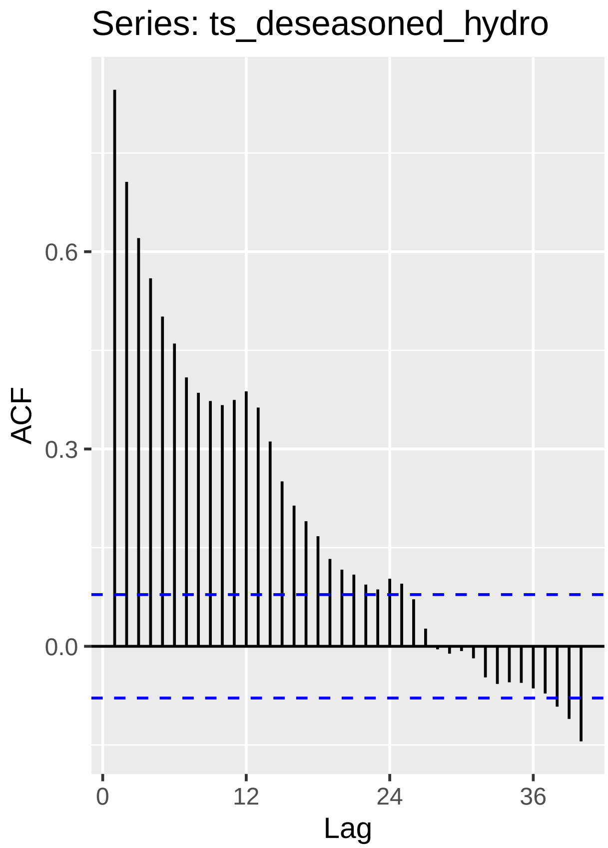
Q9

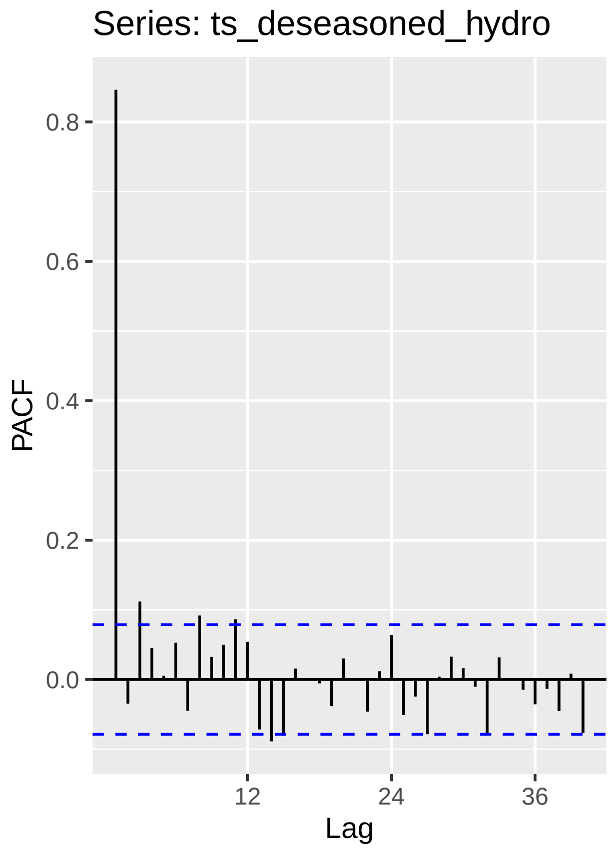
Plot ACF and PACF for the deseason series and compare with the plots from Q1. You may use plot_grid() again to get them side by side, but not mandatory. Did the plots change? How?

```r
#plot ACF
DeseasonedRenewable_acf=ggAcf(ts_deseasoned_renewable, lag.max=40, type="correlation", plot=TRUE)

DeseasonedHydroelectric_acf=ggAcf(ts_deseasoned_hydro,lag.max=40, type="correlation", plot=TRUE)

#plot PACF
DeseasonedRenewable_pacf=ggPacf(ts_deseasoned_renewable,lag.max=40, plot=TRUE)

DeseasonedHydroelectric_pacf=ggPacf(ts_deseasoned_hydro,lag.max=40, plot=TRUE)

#form into grid
plot_grid(DeseasonedRenewable_acf, DeseasonedRenewable_pacf, ncol = 2)
```




```
plot_grid(DeseasonedHydroelectric_acf, DeseasonedHydroelectric_pacf, ncol = 2)
```





For renewable production, the ACF plot didn't show much of a change since there isn't seasonality. The PACF plot has that similar strong spike at lag 1, but then unlike Q1, there aren't spikes at months 13, 25, and 27 since there seasonality has been removed.

For hydro, the ACF plot looks drastically different in Q1 vs Q9. In Q1, there was a seasonal pattern of spikes throughout the lags, while in Q9 there is a downwards trend. This change is because hydro is seasonal so when seasonality was removed, the ACF plot changed since the correlation between the data changed without seasonality. PACF also changed with fewer significant spikes, and this is due to the same reason of the seasonality being removed.