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

Assignment 3 - Due date 02/04/25

Justin Maynard

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.\

```
rm(list=ls())
#Load/install required package here
library(forecast)
library(tseries)
library(dplyr)
library(here)
library(readxl)
library(ggplot2)
library(cowplot)
```

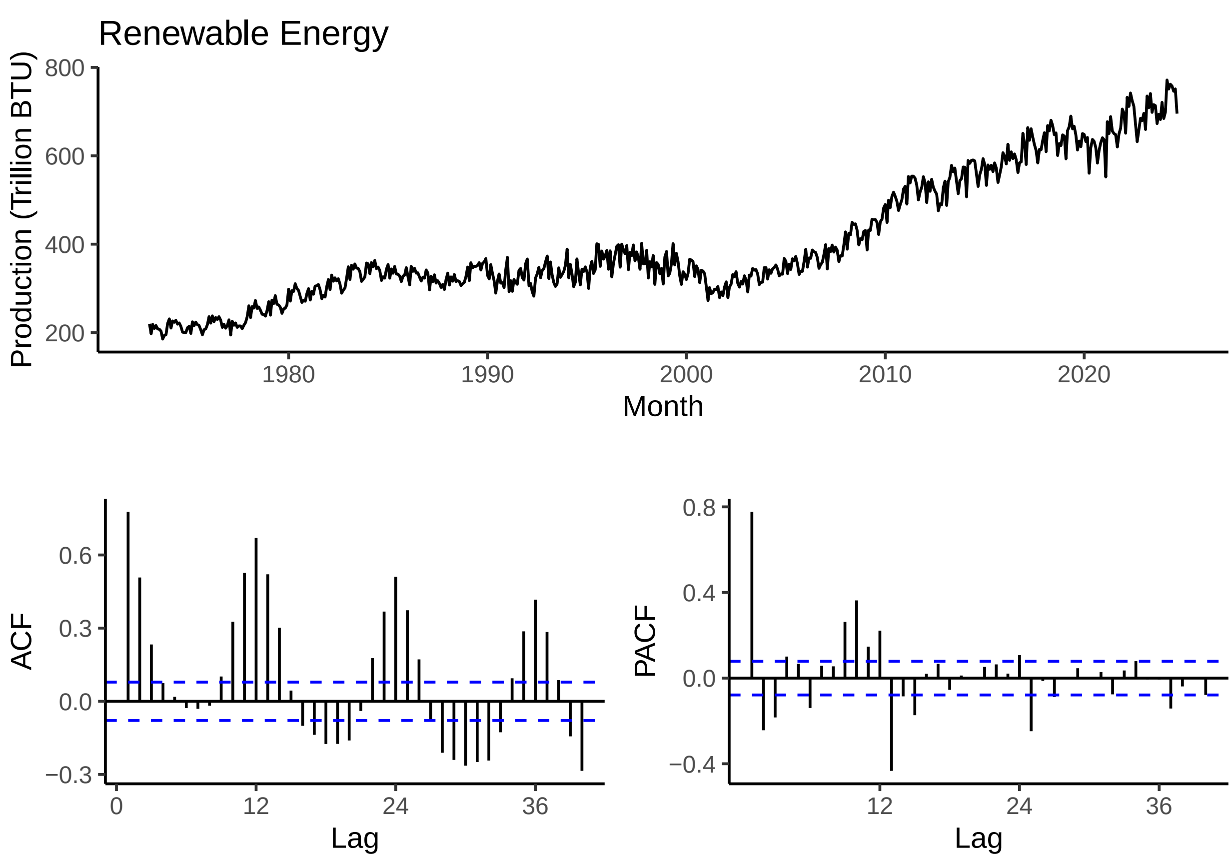
```r
data <- read_excel("Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source.xlsx",
                   skip = 12, col_names = FALSE)
names <- read_excel("Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source.xlsx",
                    skip = 10, n_max = 1, col_names = F)
colnames(data) <- names

data <- data[,c("Month", "Total Renewable Energy Production", "Hydroelectric Power Consumption")]
```

##Trend Component

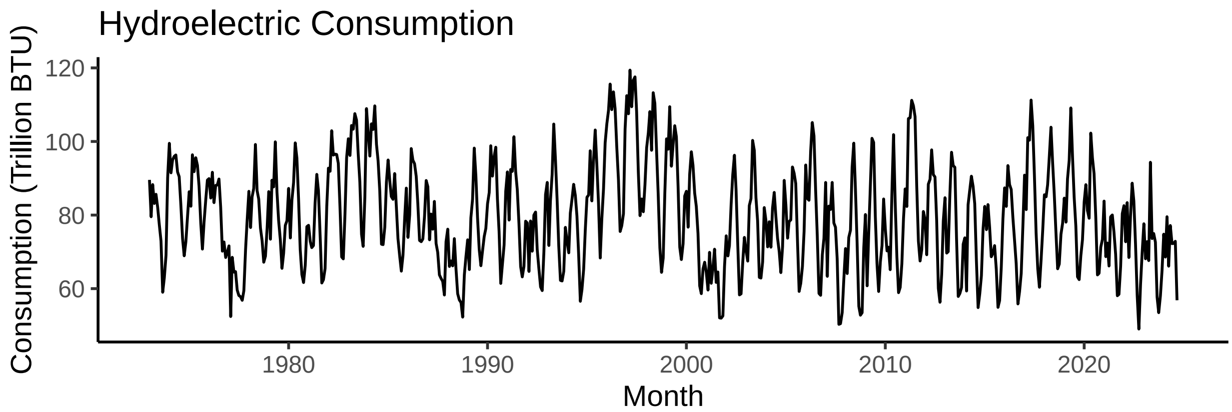
Q1

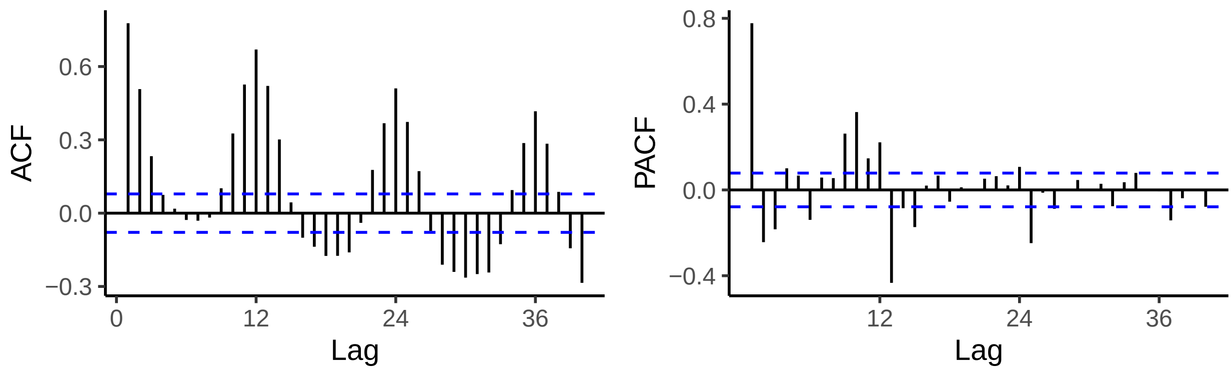
For each time series, i.e., Renewable Energy Production and Hydroelectric Consumption create three plots: one with time series, one with the ACF and with the PACF. You may use the some code form A2, but I want all the three plots side by side as in a grid. (Hint: use function `plot_grid()` from the `cowplot` package)

```r
print(plot_grid(ts_graph_renew, bottom_rewew, ncol = 1 ))
```




```r
plot_grid(ts_graph_hydro, bottom_rewew, 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?

Renewable energy production appears to have an increasing trend as production tends to increase over time. Hydroelectric power consumption appears to have no to little trend as consumption varies but does not increase or decrease over time.

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.

```
nobs <- nrow(data)
t <- c(1:nobs)

linear_trend_model_renew <- lm(ts[,1] ~ t)

print(summary(linear_trend_model_renew))
```

```
##
## Call:
```

```
## lm(formula = ts[, 1] ~ t)
##
## 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
```

```
intercept_renew <- linear_trend_model_renew$coefficients[1]
t_renew <- linear_trend_model_renew$coefficients[2]
```

```
linear_trend_model_hydro <- lm(ts[,2] ~ t) # + 1 refers to date column

print(summary(linear_trend_model_hydro))
```

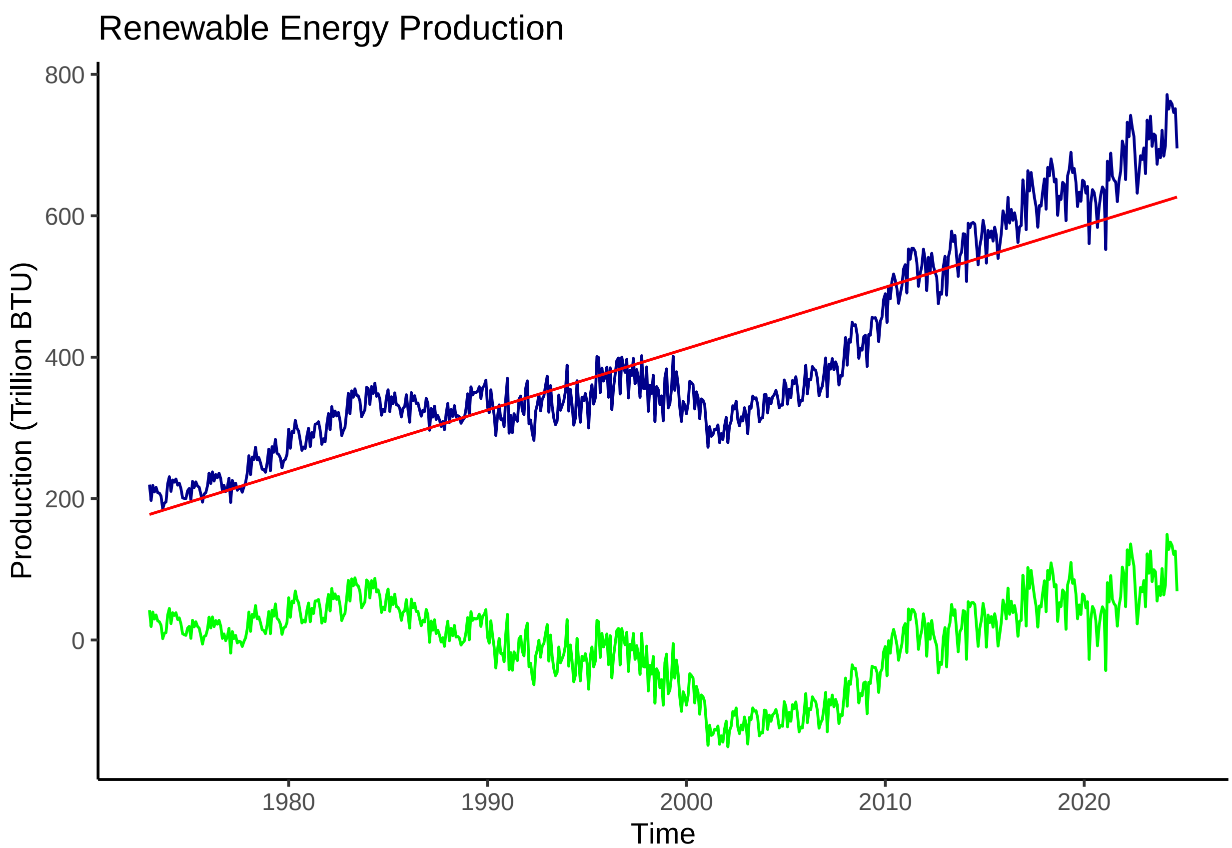
```
##
## Call:
## lm(formula = ts[, 2] ~ t)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -29.995 -10.422  -0.720   9.161  39.624
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 82.96766    1.12339  73.855  < 2e-16 ***
## t           -0.01098    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
```

```
intercept_hydro <- linear_trend_model_hydro$coefficients[1]
t_hydro <- linear_trend_model_hydro$coefficients[2]
```

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?

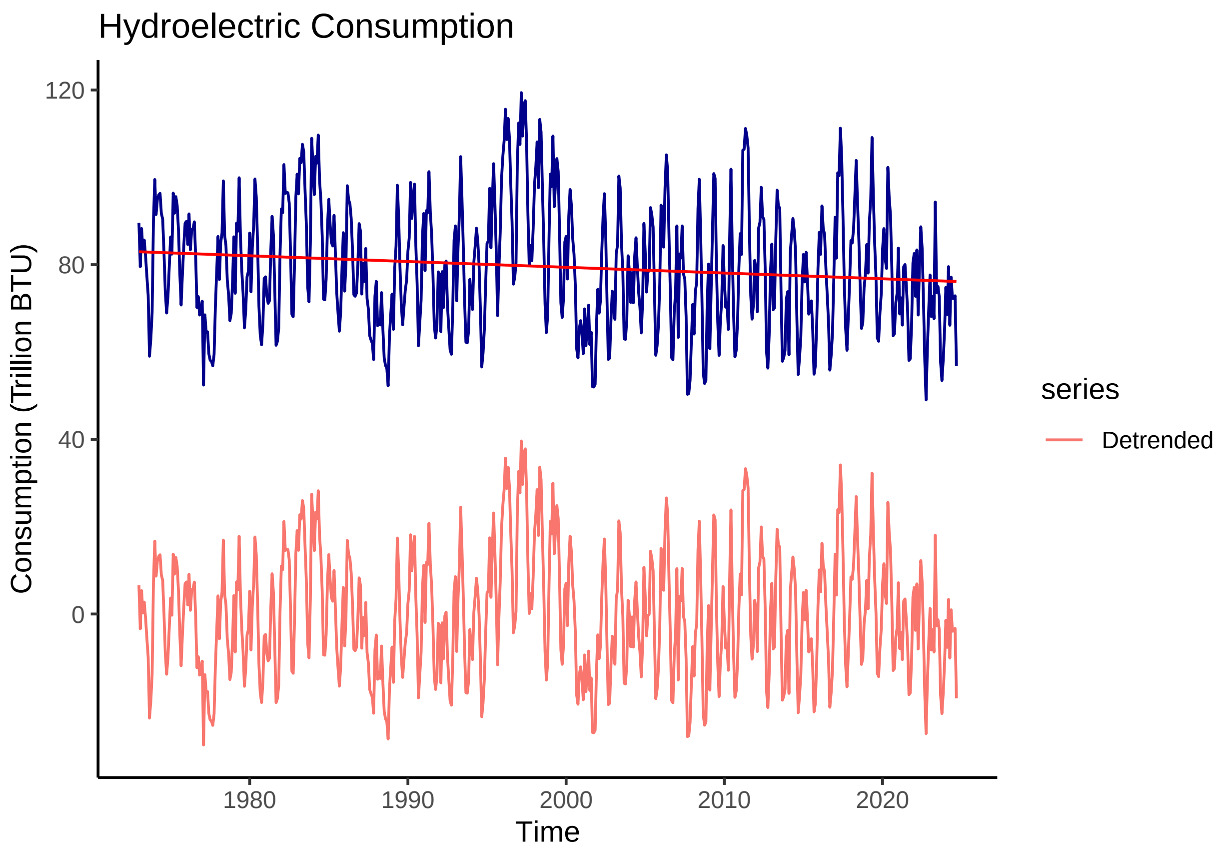
```r
linear_trend_renew <- intercept_renew + t_renew * t
ts_linear_renew <- ts(linear_trend_renew,start=c(1973,1,1),frequency=12)

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

#Plot
autoplot(ts[,1],color="darkblue")+
  autolayer(ts_detrend_renew,series="Detrended",color="green")+
  autolayer(ts_linear_renew,series="Linear Component",color="red") +
  ylab("Production (Trillion BTU)") +
  ggtitle("Renewable Energy Production") +
  theme_classic()
```



When renewable energy production is detrended the time series no longer increases across time with a linear trend, and instead just has seasonal variation. The series is centered on zero.

```r
linear_trend_hydro <- intercept_hydro + t_hydro * t
ts_linear_hydro <- ts(linear_trend_hydro,start=c(1973,1,1),frequency=12)

detrend_hydro <- ts[,2] - linear_trend_hydro
ts_detrend_hydro <- ts(detrend_hydro, start = c(1973,1,1),frequency = 12)

#Plot
autoplot(ts[,2],color="darkblue")+
  autolayer(ts_detrend_hydro,series="Detrended")+
```

```
  autolayer(ts_linear_hydro,series="Linear Component",color="red") +
  ylab("Consumption (Trillion BTU)") +
  ggtitle("Hydroelectric Consumption") +
  theme_classic()
```

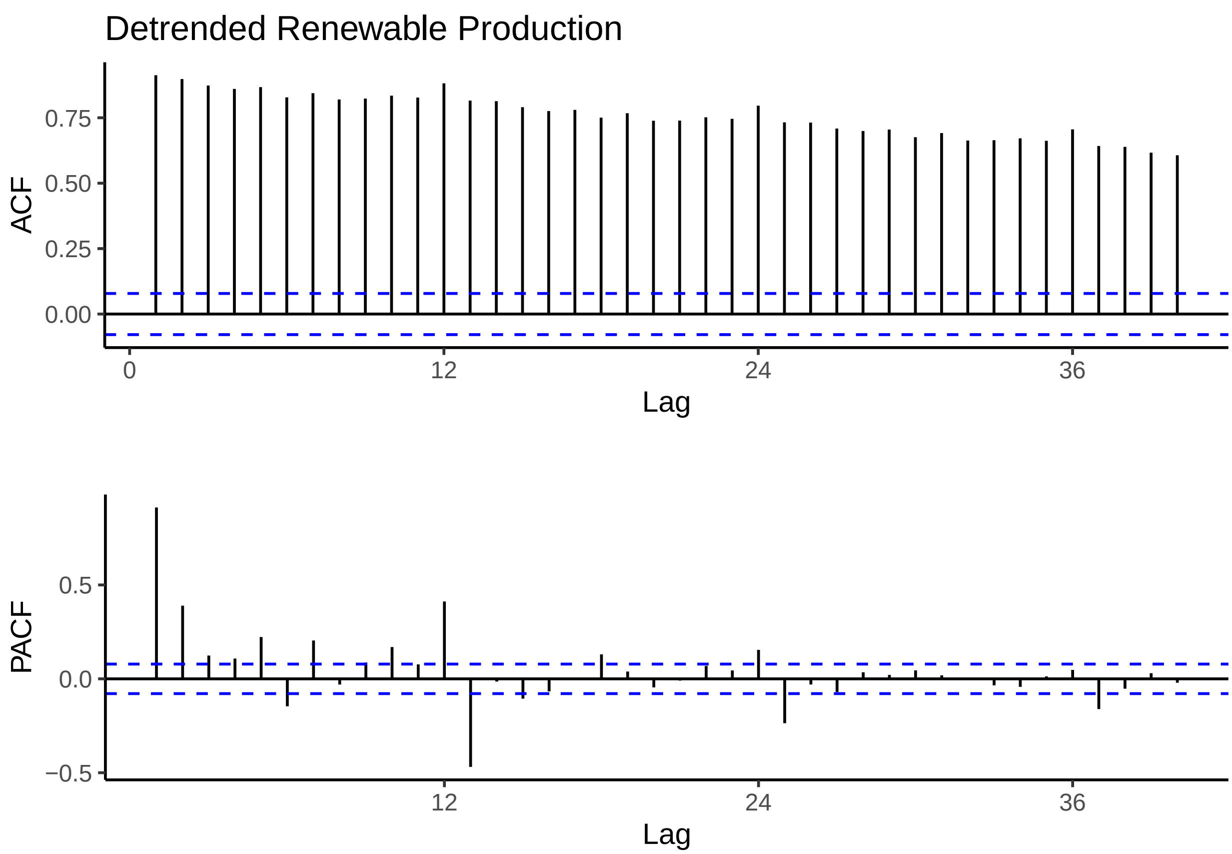


When hydroelectric consumption is detrended, the slight negative trend is removed and consumption is centered on a mean of zero and exhibits seasonal variation.

Q5

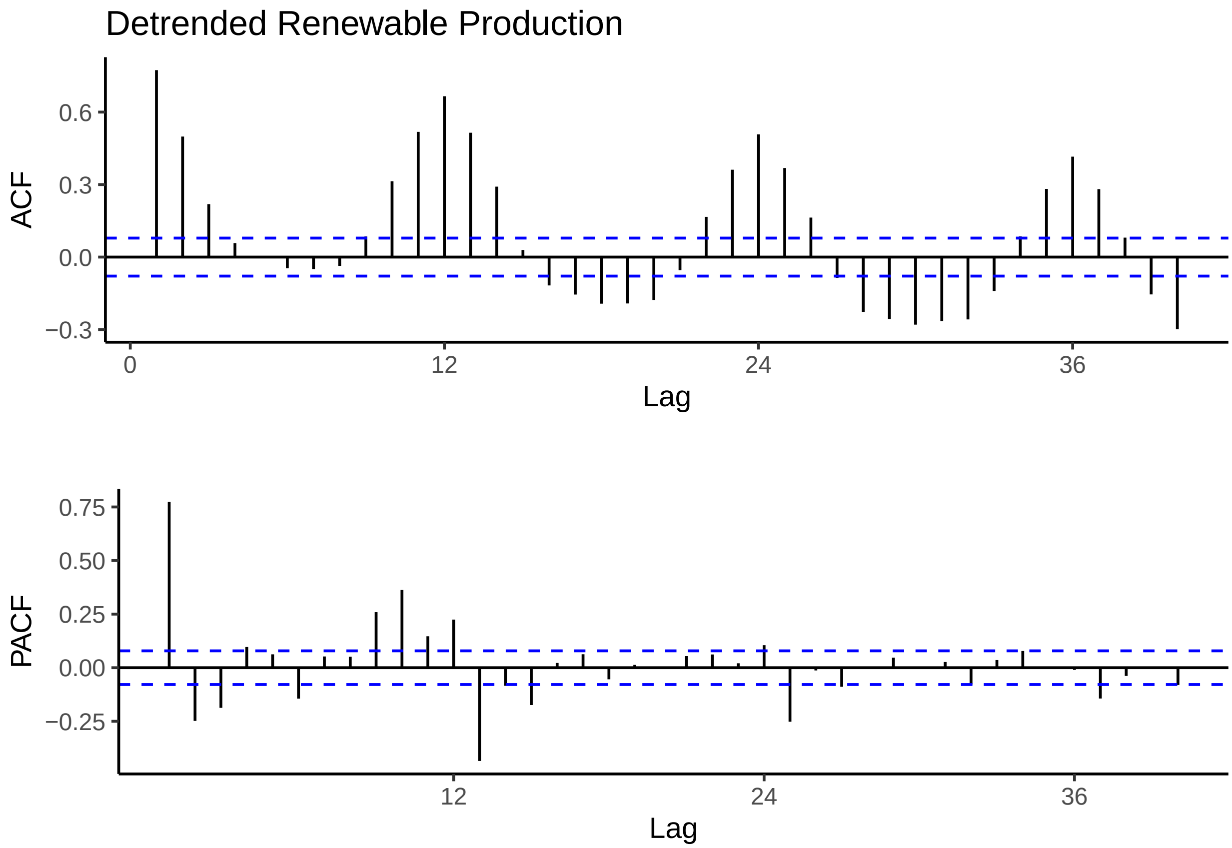
Plot ACF and PACF for the detrended series and compare with the plots from Q1. You may use plot_grid() again to get them side by side, but not mandatory. Did the plots change? How?

```
plot_grid(autoplot(acf_graph_renew_detrend)+theme_classic() + ggtitle("Detrended Renewable Production")
```



The ACF plot changes as it changed from a pattern of decreasing and increasing in a wave like shape to steadily but slowly decreasing. This is because without trend current values are more correlated with past values. PACF changed to show strong partial autocorrelation during lag 1,13,25, and 37 with a generally decreasing PACF other than those months.

```
plot_grid(autoplot(acf_graph_hydro_detrend)+theme_classic() + ggtitle("Detrended Renewable Production")
```



The detrended ACF shows many similiarities to the previous ACF, with the cyclical ACF. It appears that the cyclical nature is getting larger in the detrended ACF. The PACF values also appear to be larger in the detrened plot, and follow the same general pattern.

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.

They both appear to have seasonal trends, with seasonal variations. Hydroelectric consumption seems to have larger seasonal variations, and renewable energy production may have seasonal impacts from only certain months.

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?

```r
dummies_renew <- seasonaldummy(ts_detrend_renew)

seas_means_model_renew <- lm(detrend_renew ~ dummies_renew)
summary(seas_means_model_renew)
```

```
##
## Call:
## lm(formula = detrend_renew ~ dummies_renew)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -149.18  -38.16   14.42   41.50  134.67
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)         7.858      8.504   0.924  0.35584
## dummies_renewJan    5.592     11.968   0.467  0.64048
## dummies_renewFeb  -31.452     11.968  -2.628  0.00881 **
## dummies_renewMar    6.892     11.968   0.576  0.56491
## dummies_renewApr   -6.449     11.968  -0.539  0.59023
## dummies_renewMay    7.923     11.968   0.662  0.50822
## dummies_renewJun   -3.394     11.968  -0.284  0.77682
## dummies_renewJul    2.126     11.968   0.178  0.85906
## dummies_renewAug   -5.878     11.968  -0.491  0.62351
## dummies_renewSep  -31.209     11.968  -2.608  0.00934 **
## dummies_renewOct  -18.757     12.026  -1.560  0.11937
## dummies_renewNov  -19.982     12.026  -1.661  0.09713 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 60.73 on 609 degrees of freedom
## Multiple R-squared:  0.04839,    Adjusted R-squared:  0.0312
## F-statistic: 2.815 on 11 and 609 DF,  p-value: 0.001358
```

Renewable energy production has a statistically significant trend during the months of Feburary and September.

```r
dummies_hydro <- seasonaldummy(ts_detrend_renew)

seas_means_model_hydro <- lm(detrend_hydro ~ dummies_hydro)
summary(seas_means_model_hydro)
```

```
##
## Call:
## lm(formula = detrend_hydro ~ dummies_hydro)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -33.933  -5.798  -0.531   5.721  32.166
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
```

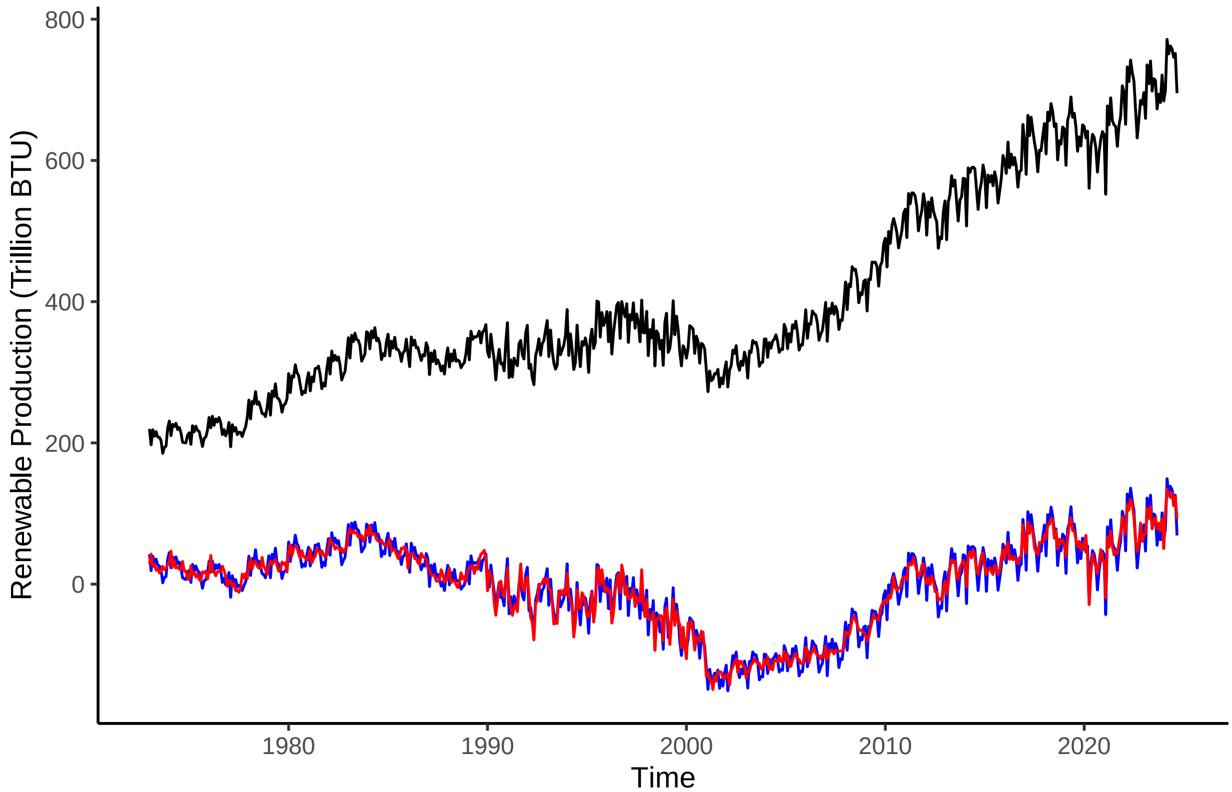
```
## (Intercept)        0.4379     1.4258   0.307 0.758849
## dummies_hydroJan   4.8863     2.0067   2.435 0.015177 *
## dummies_hydroFeb  -2.5567     2.0067  -1.274 0.203116
## dummies_hydroMar   7.0202     2.0067   3.498 0.000502 ***
## dummies_hydroApr   5.3770     2.0067   2.680 0.007572 **
## dummies_hydroMay  13.8957     2.0067   6.925 1.11e-11 ***
## dummies_hydroJun  10.7293     2.0067   5.347 1.27e-07 ***
## dummies_hydroJul   4.0439     2.0067   2.015 0.044320 *
## dummies_hydroAug  -5.3775     2.0067  -2.680 0.007566 **
## dummies_hydroSep -16.5635     2.0067  -8.254 9.51e-16 ***
## dummies_hydroOct -16.3915     2.0164  -8.129 2.43e-15 ***
## dummies_hydroNov -10.8163     2.0164  -5.364 1.16e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.18 on 609 degrees of freedom
## Multiple R-squared:  0.4781, Adjusted R-squared:  0.4687
## F-statistic: 50.72 on 11 and 609 DF,  p-value: < 2.2e-16
```

Hydroelectric consumption has statistically significant seasonal variation during all months except December and Feburary. This confirms the hypothesis from Q6 that hydroelectric consumption varies significantly based on season.

Q8

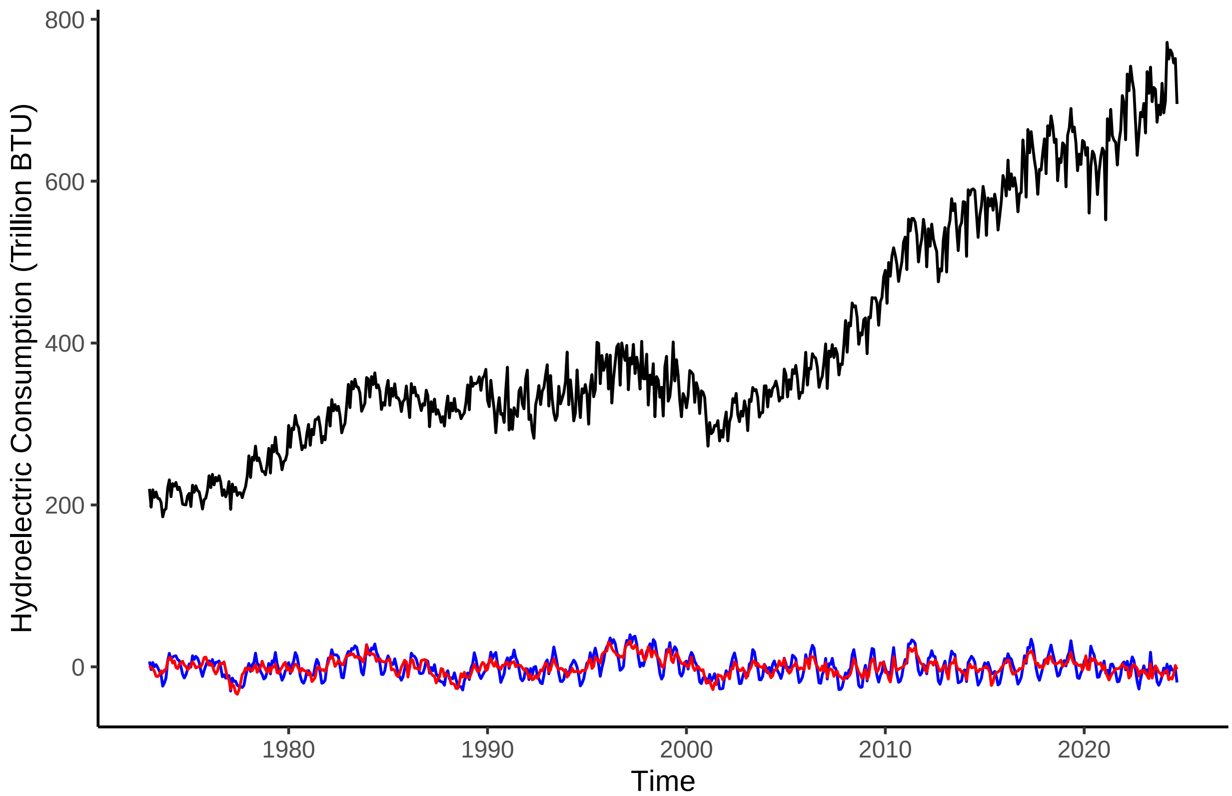
Use the regression coefficients from Q7 to deseason the series. Plot the deseason series and compare with the plots from part Q1. Did anything change?

```r
beta_intercept_renew <-seas_means_model_renew$coefficients[1]
beta_coeff_renew <-seas_means_model_renew$coefficients[2:12]

renew_seas_comp <- array(0,nobs)
for(i in 1:nobs){
  renew_seas_comp[i] <- beta_intercept_renew + beta_coeff_renew %*% dummies_renew[i,]
}


deseason_renewables <- detrend_renew - renew_seas_comp
ts_deseson_renew <- ts(deseason_renewables, start = c(1973,1), frequency = 12)

autoplot(ts[,1])+
  autolayer(ts_detrend_renew, color = "blue") +
  autolayer(ts_deseson_renew,color="red") +
  ylab("Renewable Production (Trillion BTU)") +
  theme_classic()
```



After plotting the deseasoned series (in red) some seasonal variations were smoothed with less peaks, but the series follows the same shape as the detrended series.

```
beta_intercept_hydro <-seas_means_model_hydro$coefficients[1]
beta_coeff_hydro <-seas_means_model_hydro$coefficients[2:12]

hydro_seas_comp <- array(0,nobs)
for(i in 1:nobs){
  hydro_seas_comp[i] <- beta_intercept_hydro + beta_coeff_hydro %*% dummies_hydro[i,]
}


deseason_hydro <- detrend_hydro - hydro_seas_comp
ts_deseson_hydro <- ts(deseason_hydro, start = c(1973,1), frequency = 12)

autoplot(ts[,1])+
  autolayer(ts_detrend_hydro, color = "blue", series = "Detrended") +
  autolayer(ts_deseson_hydro,color="red", series = "Deseasoned") +
  labs(color = "Trend") +
  ylab("Hydroelectric Consumption (Trillion BTU)") +
  theme_classic()
```

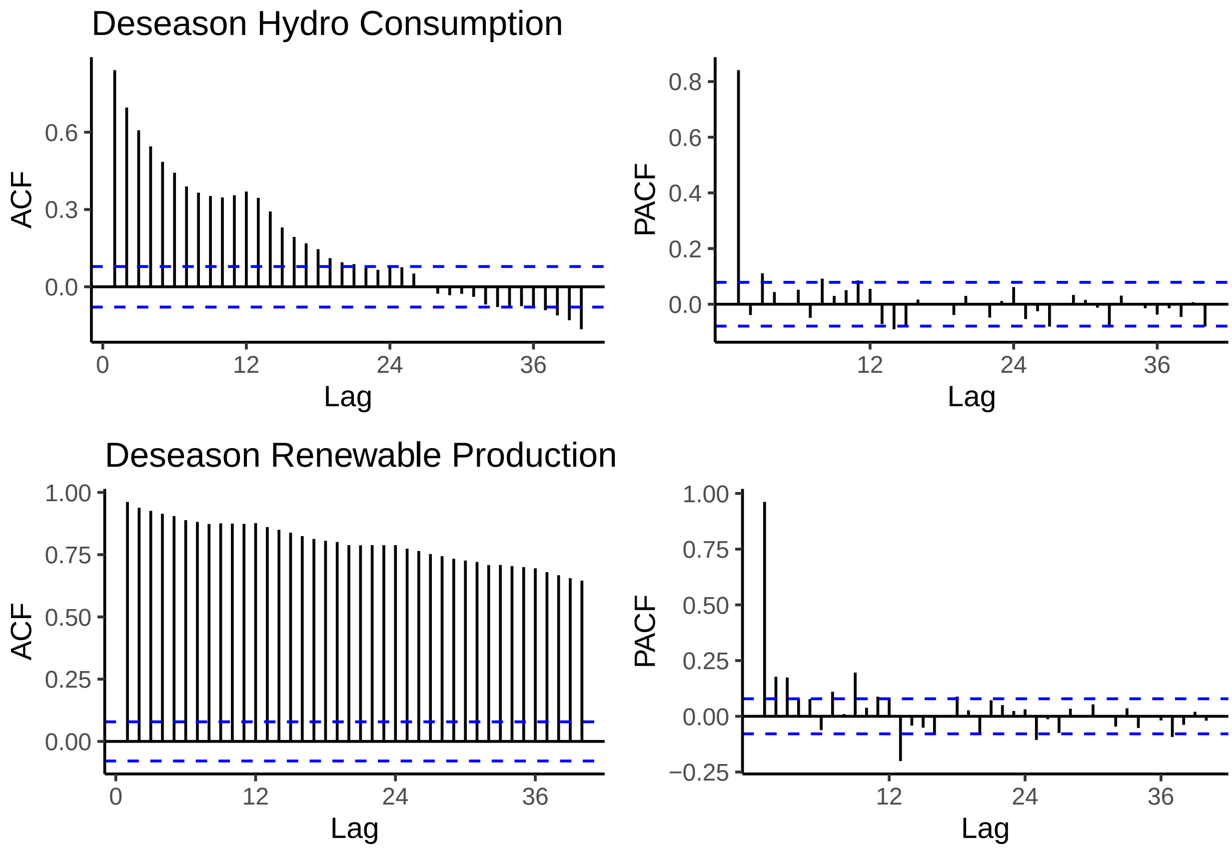


The deseasoned time series has a bigger smoothing effect for the hydroelectric consumption, as there are more significant months with seasonality. The seasonal swings are eliminated from the time series.

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?

```
plot_grid(autoplot(acf_graph_hydro_deseason)+theme_classic() + ggtitle("Deseason Hydro Consumption"), au
          autoplot(acf_graph_renew_deseason) + theme_classic() + ggtitle("Deseason Renewable Production
          autoplot(pacf_graph_renew_deseason) + theme_classic() + ggtitle("")
          )
```



Compared to Q1 the ACF of the renewable energy production is much different, with decreasing ACFs that start at a much higher value and only decrease to ~.75, while the maximum value of the original ACF was ~.75. The PACF is also different, with a much higher initial PACF and then many statistically significantly different PACFs. The ACF of the hydro consumption is also much different, and decreases quickly to significant ACFs. The PACF follows a similar pattern to the renewable produciton, with a high initial value and then significant PACFs following that.