ENV 790.30 - Time Series Analysis for Energy Data | Spring 2024

Assignment 3 - Due date 02/01/24

Cara Kuuskvere

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_A02_Sp24.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 2022 **Monthly** Energy Review. Once again you will work only with the following columns: Total Biomass Energy Production, Total Renewable Energy Production, Hydroelectric Power Consumption. Create a data frame structure with these three 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(readr)
library(ggplot2)
library(cowplot)
```

```r
#Importing data set
raw_energy_data <- read.csv(
  "Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source.csv",
  header=FALSE,skip=12)
```

```r
#trim the table
energy_data <- raw_energy_data[,4:6] #want all rows
                                  #all columns from 4 to 6
n_energy_sources <- ncol(energy_data) #number of variables
n_obs <- nrow(energy_data) #number observations

#Adding column names
colnames(energy_data)=c("Total Biomass Energy Production (Trillion Btu)",
                        "Total Renewable Energy Production (Trillion Btu)",
                        "Hydroelectric Power Consumption (Trillion Btu)")

head(energy_data)
```

```
##   Total Biomass Energy Production (Trillion Btu)
## 1                                         129.824
## 2                                         130.807
## 3                                         118.091
## 4                                         130.727
## 5                                         126.583
## 6                                         130.789
##   Total Renewable Energy Production (Trillion Btu)
## 1                                           220.755
## 2                                           231.010
## 3                                           210.188
## 4                                           226.384
## 5                                           223.218
## 6                                           227.793
##   Hydroelectric Power Consumption (Trillion Btu)
## 1                                          90.131
## 2                                          99.500
## 3                                          91.476
## 4                                          94.950
## 5                                          95.969
## 6                                          96.337
```

```r
ts_energy_data <- ts(energy_data,start=c(1973,1),frequency = 12)
# frequency is 12 because it's monthly data, and months repeat every 12 entries!
# starts in year 1973
head(ts_energy_data)
```

```
##          Total Biomass Energy Production (Trillion Btu)
## Jan 1973                                        129.824
## Feb 1973                                        130.807
## Mar 1973                                        118.091
## Apr 1973                                        130.727
## May 1973                                        126.583
## Jun 1973                                        130.789
```

```
##          Total Renewable Energy Production (Trillion Btu)
## Jan 1973                                          220.755
## Feb 1973                                          231.010
## Mar 1973                                          210.188
## Apr 1973                                          226.384
## May 1973                                          223.218
## Jun 1973                                          227.793
##          Hydroelectric Power Consumption (Trillion Btu)
## Jan 1973                                         90.131
## Feb 1973                                         99.500
## Mar 1973                                         91.476
## Apr 1973                                         94.950
## May 1973                                         95.969
## Jun 1973                                         96.337
```

##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)

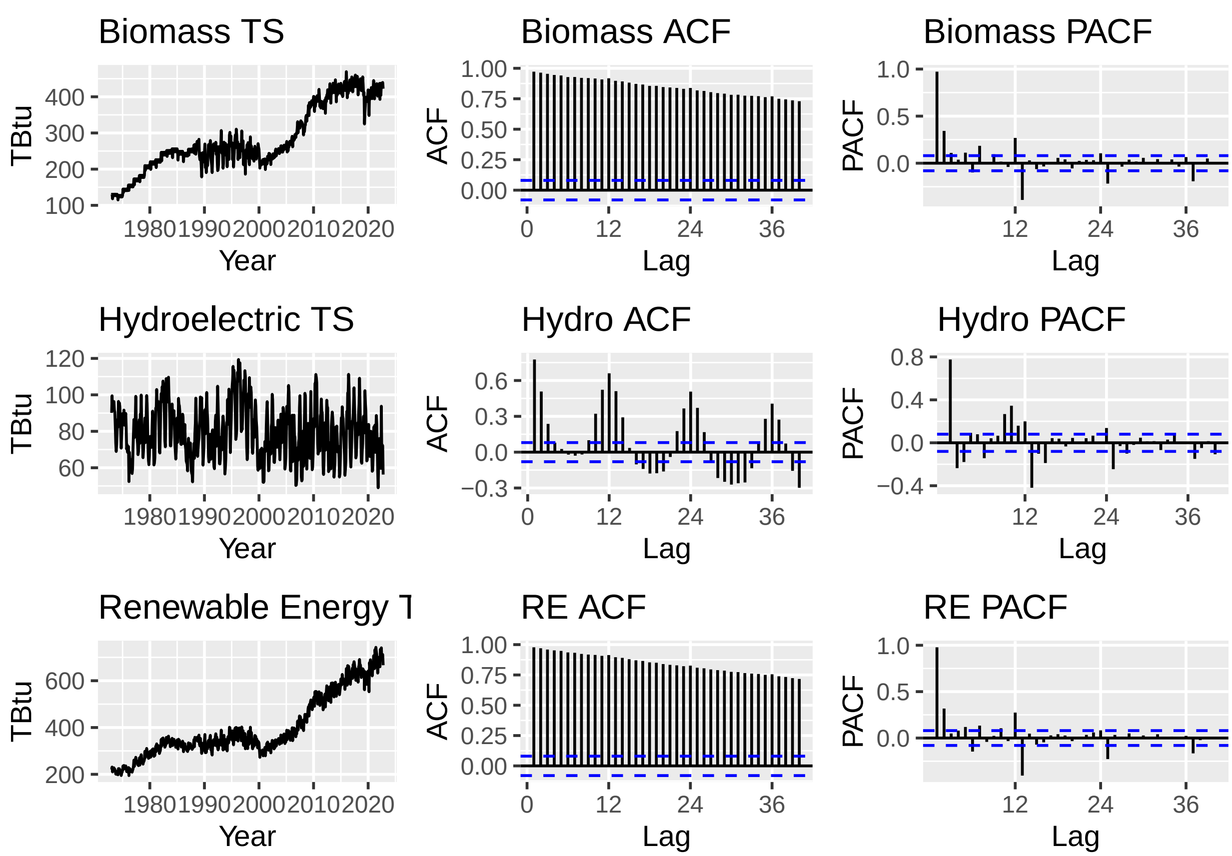
```
biomass_ts_plt <- autoplot(ts_energy_data[,1]) +
  labs(x="Year",y= "TBtu",
    title = "Biomass TS")
#biomass_ts_plt

RE_ts_plt <- autoplot(ts_energy_data[,2]) +
  labs(x="Year",y= "TBtu",
  title = "Renewable Energy TS")
#RE_ts_plt

hydro_ts_plt <- autoplot(ts_energy_data[,3]) +
  labs(x="Year",y= "TBtu",
    title = "Hydroelectric TS")
#hydro_ts_plt
```

```
biomass_acf <- ggAcf(ts_energy_data[,1],lag.max=40)+
  labs(title="Biomass ACF")
#biomass_acf
hydro_acf <- ggAcf(ts_energy_data[,3],lag.max=40)+
  labs(title="Hydro ACF")
#hydro_acf
RE_acf <- ggAcf(ts_energy_data[,2],lag.max=40)+
  labs(title="RE ACF")
#RE_acf
```

```
biomass_pacf <- ggPacf(ts_energy_data[,1],lag.max=40)+
  labs(title="Biomass PACF")
#biomass_pacf
hydro_pacf <- ggPacf(ts_energy_data[,3],lag.max=40)+
```

```
  labs(title="Hydro PACF")
#hydro_pacf
RE_pacf <- ggPacf(ts_energy_data[,2],lag.max=40)+
  labs(title="RE PACF")
#RE_pacf
```

```
#plot_grid(biomass_ts_plt,biomass_acf,biomass_pacf, align="h",label_size=8)
#plot_grid(hydro_ts_plt,hydro_acf,hydro_pacf, align = "h", label_size=8)
#plot_grid(RE_ts_plt,RE_acf,RE_pacf,align="h",label_size=8)
plot_grid(biomass_ts_plt,biomass_acf,biomass_pacf,
          hydro_ts_plt,hydro_acf,hydro_pacf,
          RE_ts_plt,RE_acf,RE_pacf,
          ncol=3,align="h")
```



Q2

From the plot in Q1, do the series Total Biomass Energy Production, Total Renewable Energy Production, Hydroelectric Power Consumption appear to have a trend? If yes, what kind of trend?

Biomass and renewable energy production appear to have an increasing somewhat linear trend over time. The hydroelectric power consumption appears to have a strong seasonal component, which makes it hard to see if there is a small trend component.

Q3

Use the *lm()* function to fit a linear trend to the three 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 <- 1:n_obs
biomass_linear_trend <- lm(energy_data[,1]~t)
RE_linear_trend <- lm(energy_data[,2]~t)
hydro_linear_trend <- lm(energy_data[,3]~t)

summary(biomass_linear_trend)
```

```
##
## Call:
## lm(formula = energy_data[, 1] ~ t)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -102.453  -24.240    6.106   32.219   83.377
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 140.26864    3.24250   43.26   <2e-16 ***
## t             0.47527    0.00938   50.67   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 39.6 on 596 degrees of freedom
## Multiple R-squared:  0.8116, Adjusted R-squared:  0.8113
## F-statistic:  2567 on 1 and 596 DF,  p-value: < 2.2e-16
```

```
summary(RE_linear_trend)
```

```
##
## Call:
## lm(formula = energy_data[, 2] ~ t)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -148.05  -36.16   11.72   42.31  143.58
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 187.25183    4.98485   37.56   <2e-16 ***
## t             0.70777    0.01442   49.08   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 60.87 on 596 degrees of freedom
## Multiple R-squared:  0.8017, Adjusted R-squared:  0.8013
## F-statistic:  2409 on 1 and 596 DF,  p-value: < 2.2e-16
```

```r
summary(hydro_linear_trend)
```

```
##
## Call:
## lm(formula = energy_data[, 3] ~ t)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -30.180 -10.665  -0.658   9.545  39.409
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 83.02761    1.15464   71.91  < 2e-16 ***
## t           -0.01086    0.00334   -3.25  0.00122 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.1 on 596 degrees of freedom
## Multiple R-squared:  0.01741,    Adjusted R-squared:  0.01576
## F-statistic: 10.56 on 1 and 596 DF,  p-value: 0.001219
```

The biomass linear trend has an intercept value of ~140 and a slope of .47, showing a positive linear trend with a strong correlation of a very small p-value.

The renewable energy linear trend has an intercept of ~187 and a slope of ~.7, showing a positive linear trend with a strong correlation of a very small p-value.

The hydro linear trend has an intercept of ~83 with a very strong, small p value,but a small, weak, negative linear trend with slope -0.01 with a p value that is still less than 0.05. It also has a weaker r squared value than the other two trends.

Q4

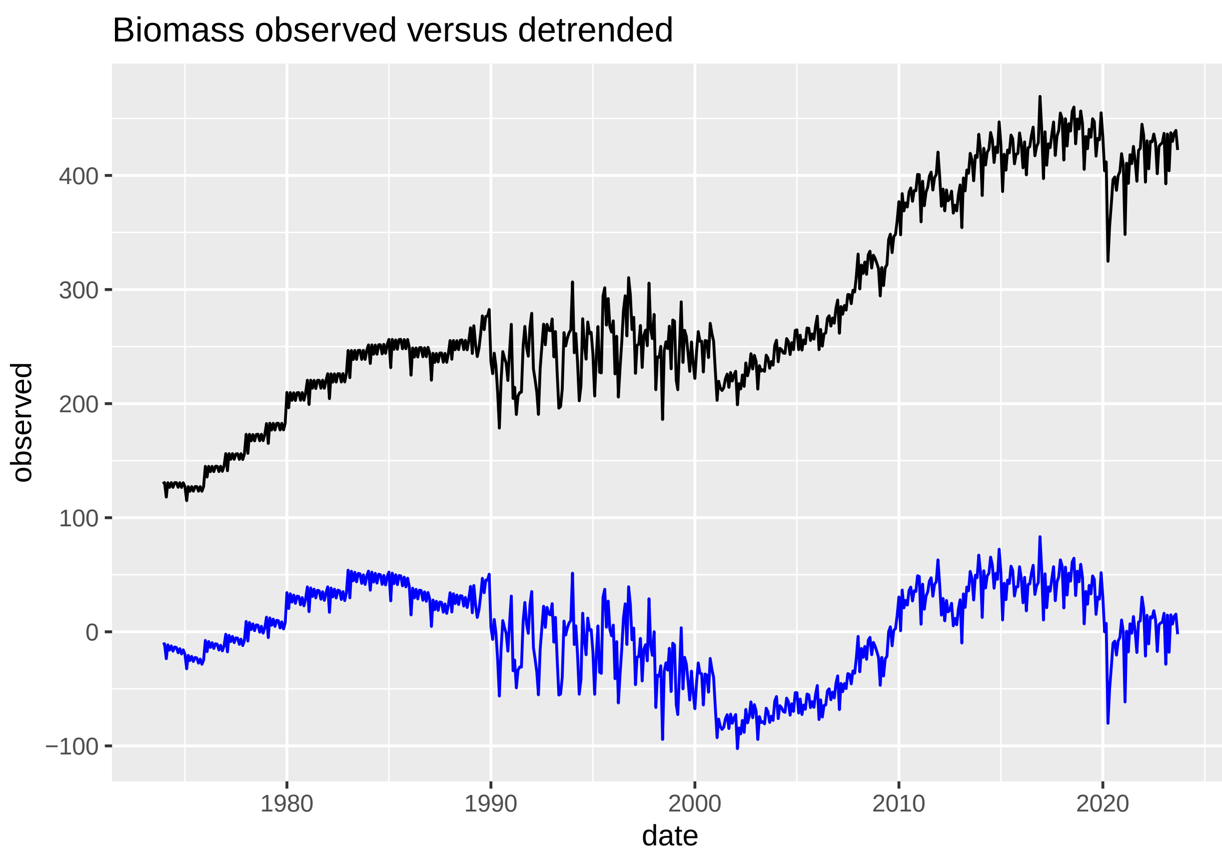
Use the regression coefficients from Q3 to detrend the series. Plot the detrended series and compare with the plots from Q1. What happened? Did anything change?

```r
library(lubridate)
```

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

## The following object is masked from 'package:cowplot':
##
##     stamp

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

```r
ym_date <- paste(raw_energy_data[,1])
ym_date <- ym(ym_date)  #function my from package lubridate
```

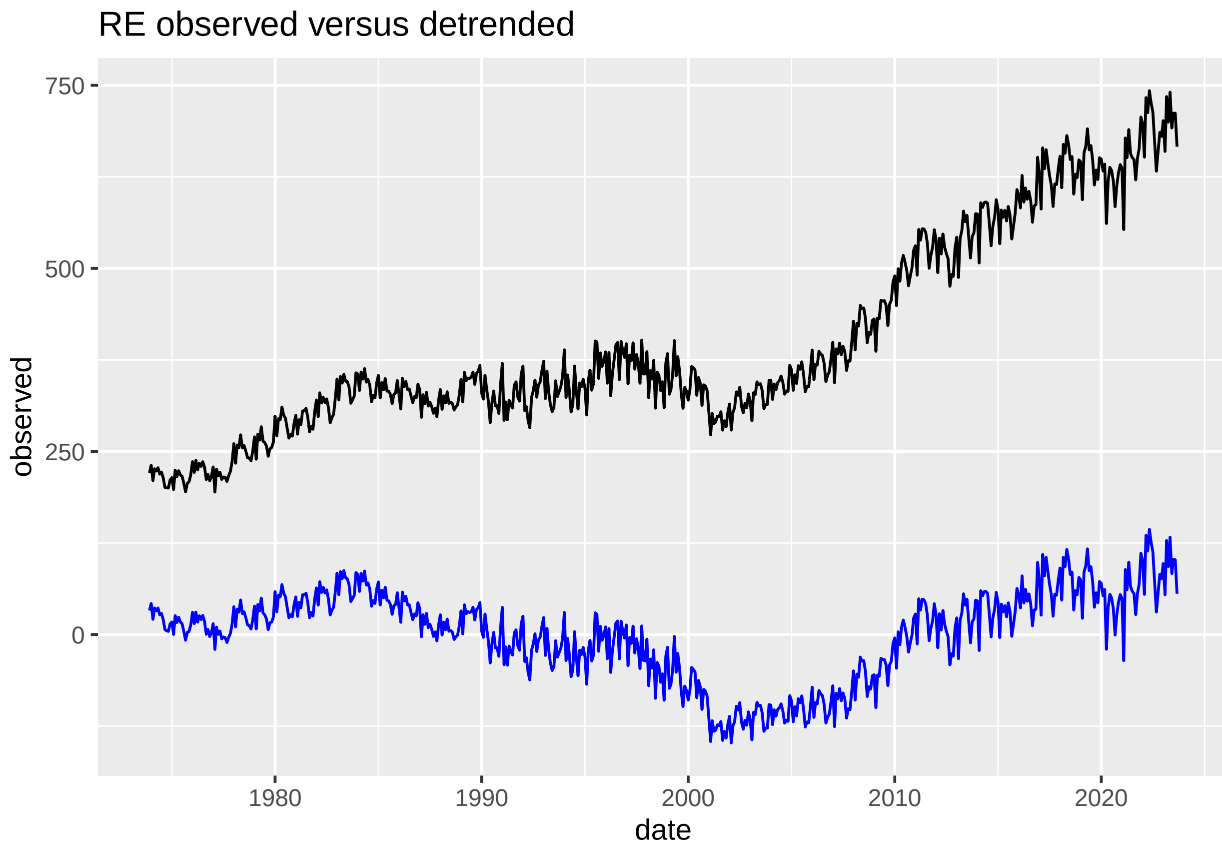
```r
biomass_beta0<- biomass_linear_trend$coefficients[1]
biomass_beta1 <- biomass_linear_trend$coefficients[2]


biomass_detrend <- energy_data[,1] - (biomass_beta0 + biomass_beta1*t)

df_biomass_detrend <-data.frame("date"=ym_date,
                        "observed"=energy_data[,1],
                        "detrend" =biomass_detrend)
ggplot(df_biomass_detrend, aes(x=date))+
  geom_line(aes(y=observed),color="black")+
  geom_line(aes(y=detrend),color="blue")+
  labs(title="Biomass observed versus detrended")
```



This plot shows the upward linear trend removed, as well as centered the mean around zero. It also shows a greater dip around the year 2000.

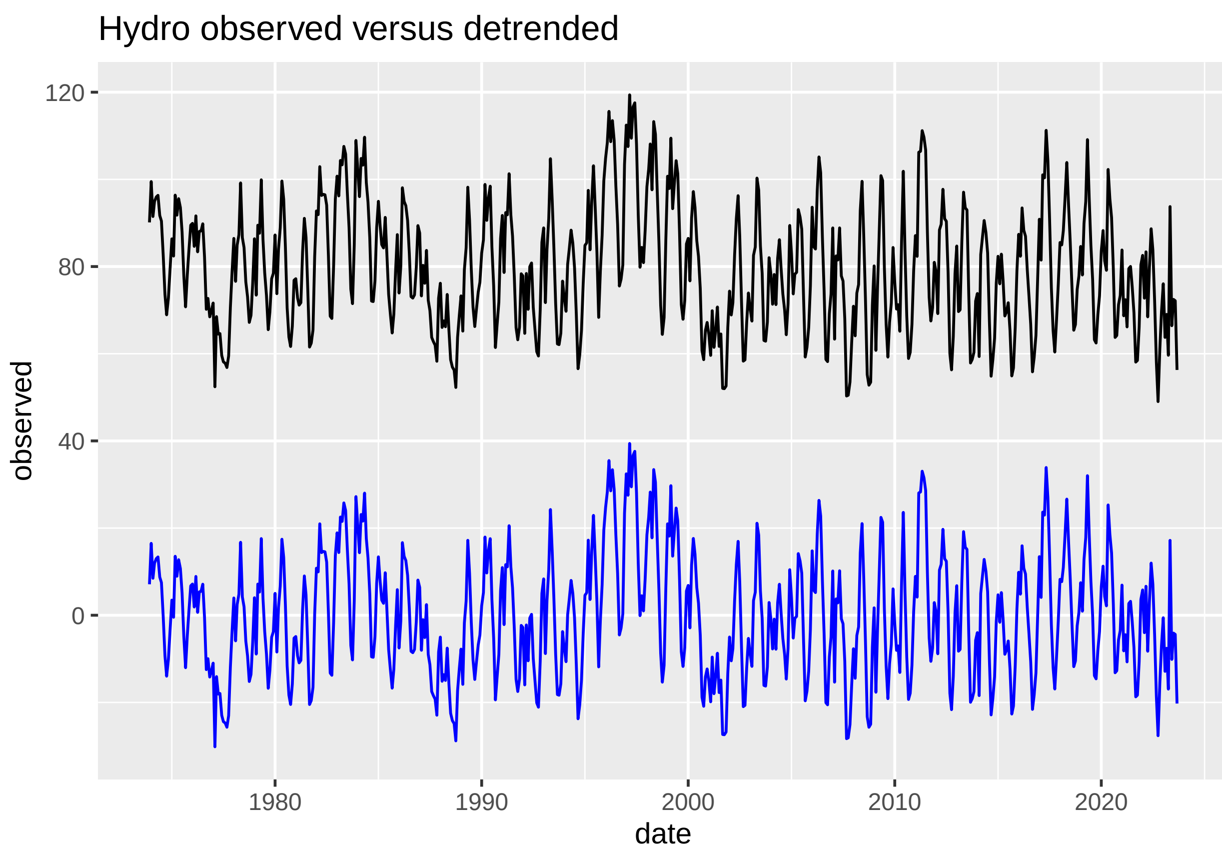
```r
RE_beta0<- RE_linear_trend$coefficients[1]
RE_beta1 <- RE_linear_trend$coefficients[2]


RE_detrend <- energy_data[,2] - (RE_beta0 + RE_beta1*t)

df_RE_detrend <-data.frame("date"=ym_date,
                        "observed"=energy_data[,2],
                        "detrend" =RE_detrend)
```

```r
ggplot(df_RE_detrend, aes(x=date))+
  geom_line(aes(y=observed),color="black")+
  geom_line(aes(y=detrend),color="blue")+
  labs(title="RE observed versus detrended")
```



This plot removed the upward linear trend was removed and centered the mean at zero. It also shows a greater dip after 2000 relative to the rest of the data.

```r
hydro_beta0<- hydro_linear_trend$coefficients[1]
hydro_beta1 <- hydro_linear_trend$coefficients[2]


hydro_detrend <- energy_data[,3] - (hydro_beta0 + hydro_beta1*t)

df_hydro_detrend <-data.frame("date"=ym_date,
                        "observed"=energy_data[,3],
                        "detrend" =hydro_detrend)
ggplot(df_hydro_detrend, aes(x=date))+
  geom_line(aes(y=observed),color="black")+
  geom_line(aes(y=detrend),color="blue")+
  labs(title="Hydro observed versus detrended")
```



This plot did not change significantly beyond centering the mean at zero because there was very little trend component to this 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. nut mot mandatory. Did the plots change? How?

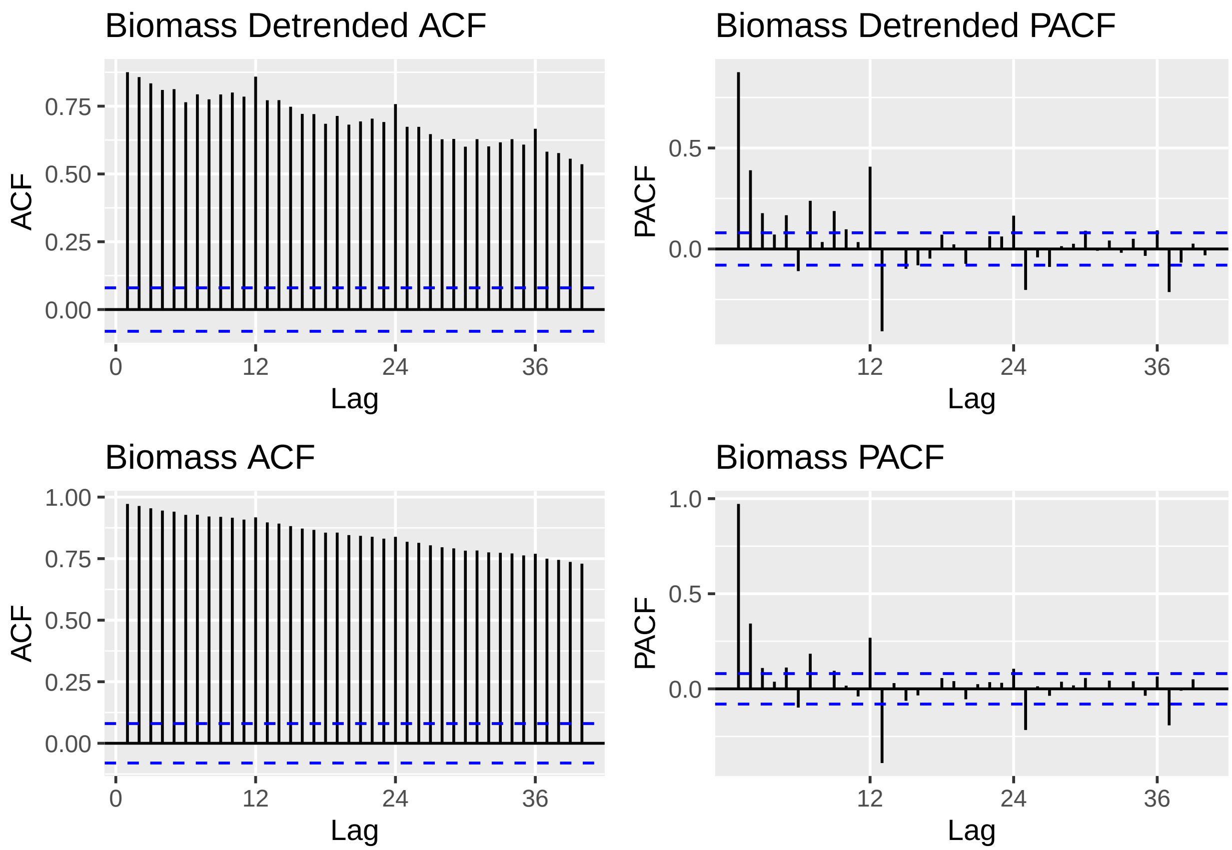
```
ts_biomass_detrend <- ts(biomass_detrend,start=c(1973,1),frequency = 12)
ts_RE_detrend <- ts(RE_detrend,start=c(1973,1),frequency = 12)
ts_hydro_detrend <- ts(hydro_detrend,start=c(1973,1),frequency = 12)
```

```
biomass_detrend_acf <- ggAcf(ts_biomass_detrend,lag.max=40)+
  labs(title="Biomass Detrended ACF")
#biomass_detrend_acf
hydro_detrend_acf <- ggAcf(ts_hydro_detrend,lag.max=40)+
  labs(title="Hydro Detrended ACF")
#hydro_detrend_acf
RE_detrend_acf <- ggAcf(ts_RE_detrend,lag.max=40)+
  labs(title="RE Detrended ACF")
#RE_detrend_acf
```

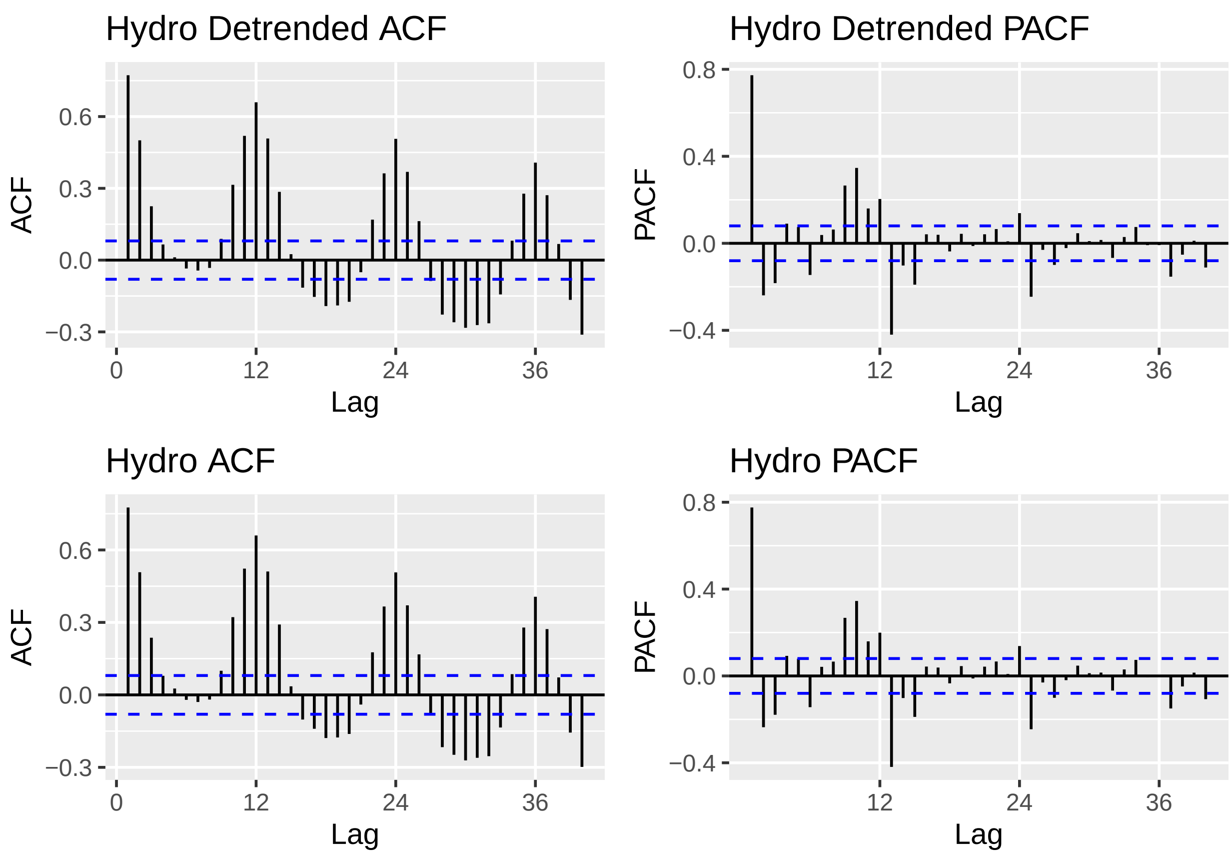
```
biomass_detrend_pacf <- ggPacf(ts_biomass_detrend,lag.max=40)+
  labs(title="Biomass Detrended PACF")
#biomass_detrend_pacf
```

```
hydro_detrend_pacf <- ggPacf(ts_hydro_detrend,lag.max=40)+
  labs(title="Hydro Detrended PACF")
#hydro_detrend_pacf
RE_detrend_pacf <- ggPacf(ts_RE_detrend,lag.max=40)+
  labs(title="RE Detrended PACF")
#RE_detrend_pacf
```

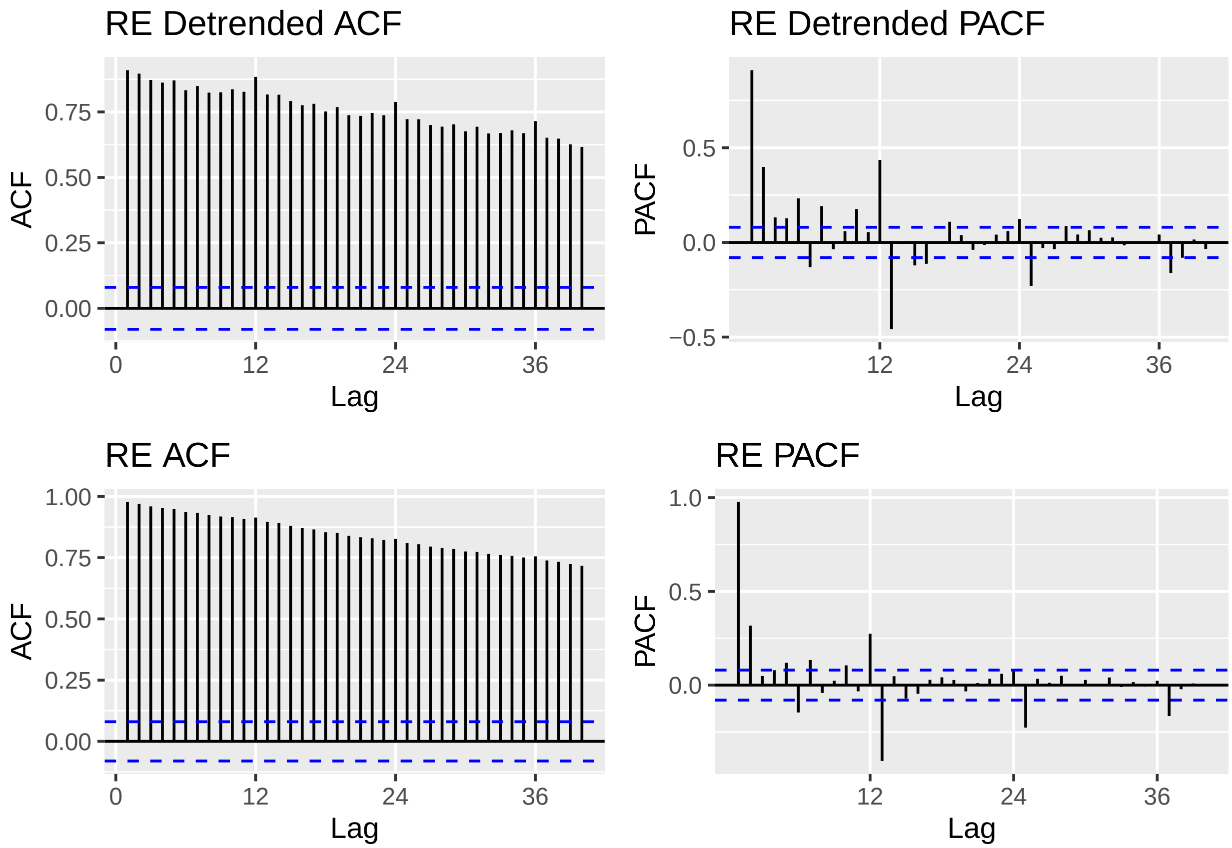
```
plot_grid(biomass_detrend_acf,biomass_detrend_pacf,biomass_acf,biomass_pacf,
          ncol=2,align="h")
```




```
plot_grid(hydro_detrend_acf,hydro_detrend_pacf ,hydro_acf,hydro_pacf,
          ncol=2,align="h")
```




```r
plot_grid(RE_detrend_acf,RE_detrend_pacf,RE_acf,RE_pacf,
          ncol=2,align="h")
```



The detrended biomass ACF shows much lower probability and connections over the periods than the detrended ACF, however the PACF shows some stronger probability on the detrended period and than on the original PACF.

The detrended hydro ACF and PACF looks very similar to the original hydro ACF and PACF because there was very little trend to remove ot begin with, so this did not make a huge impact.

The detrended RE ACF and PACF demonstrates a similar change to the detrended biomass figures.

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.

Just looking at the plots from Q1, hydro appears to have a strong seasonal trend. Biomass and RE appear to have what could be a weaker seasonal trend with their oscillating values at a somewhat regular period, with renewable energy appear to have more regular variations than biomass.

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
#Use seasonal means model

#Create the seasonal dummies
biomass_dummies <- seasonaldummy(ts_energy_data[,1])


#Fit a linear model to the seasonal dummies
biomass_seas_means_model <- lm(ts_energy_data[,1]~biomass_dummies)
summary(biomass_seas_means_model)
```

```
##
## Call:
## lm(formula = ts_energy_data[, 1] ~ biomass_dummies)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -165.85  -55.24  -28.02   96.91  178.89
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         283.9706    13.0762  21.717   <2e-16 ***
## biomass_dummiesJan    6.4960    18.3998   0.353    0.724
## biomass_dummiesFeb    8.0013    18.3998   0.435    0.664
## biomass_dummiesMar  -21.9534    18.3998  -1.193    0.233
## biomass_dummiesApr    0.6772    18.3998   0.037    0.971
## biomass_dummiesMay  -12.2418    18.3998  -0.665    0.506
## biomass_dummiesJun   -4.5013    18.3998  -0.245    0.807
## biomass_dummiesJul   -9.9153    18.3998  -0.539    0.590
## biomass_dummiesAug    6.0708    18.3998   0.330    0.742
## biomass_dummiesSep    9.2018    18.3998   0.500    0.617
## biomass_dummiesOct   -3.6518    18.3998  -0.198    0.843
## biomass_dummiesNov    5.6767    18.4925   0.307    0.759
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 91.53 on 586 degrees of freedom
## Multiple R-squared:  0.0101, Adjusted R-squared:  -0.008486
## F-statistic: 0.5433 on 11 and 586 DF,  p-value: 0.8739
```

```r
#Store regression coefficients
biomass_beta_int <- biomass_seas_means_model$coefficients[1]
biomass_beta_coeff <- biomass_seas_means_model$coefficients[2:12]
```

The biomass coefficients demonstrate there is a seasonal component, but this is a weak driver of the values in the model as the r-squared value is very low.

```r
#Use seasonal means model

#Create the seasonal dummies
RE_dummies <- seasonaldummy(ts_energy_data[,2])


#Fit a linear model to the seasonal dummies
```

```r
RE_seas_means_model <- lm(ts_energy_data[,2]~RE_dummies)
summary(RE_seas_means_model)
```

```
##
## Call:
## lm(formula = ts_energy_data[, 2] ~ RE_dummies)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -197.11  -85.79  -51.17  113.30  327.68
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    388.0347    19.6010  19.797   <2e-16 ***
## RE_dummiesJan   16.4914    27.5811   0.598    0.550
## RE_dummiesFeb   23.2065    27.5811   0.841    0.400
## RE_dummiesMar  -14.5303    27.5811  -0.527    0.599
## RE_dummiesApr   24.2156    27.5811   0.878    0.380
## RE_dummiesMay   11.5261    27.5811   0.418    0.676
## RE_dummiesJun   27.0403    27.5811   0.980    0.327
## RE_dummiesJul   16.1352    27.5811   0.585    0.559
## RE_dummiesAug   22.9346    27.5811   0.832    0.406
## RE_dummiesSep   15.3868    27.5811   0.558    0.577
## RE_dummiesOct   -8.8250    27.5811  -0.320    0.749
## RE_dummiesNov    0.3201    27.7200   0.012    0.991
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 137.2 on 586 degrees of freedom
## Multiple R-squared:  0.009313,   Adjusted R-squared:  -0.009283
## F-statistic: 0.5008 on 11 and 586 DF,  p-value: 0.9031
```

```r
#Store regression coefficients
RE_beta_int <- RE_seas_means_model$coefficients[1]
RE_beta_coeff <- RE_seas_means_model$coefficients[2:12]
```

The coefficients demonstrate that there are seasonal components that affect the renewable energy as well, but these are very week as noted by their low R squared value.

```r
#Use seasonal means model

#Create the seasonal dummies
hydro_dummies <- seasonaldummy(ts_energy_data[,3])


#Fit a linear model to the seasonal dummies
hydro_seas_means_model <- lm(ts_energy_data[,3]~hydro_dummies)
summary(hydro_seas_means_model)
```

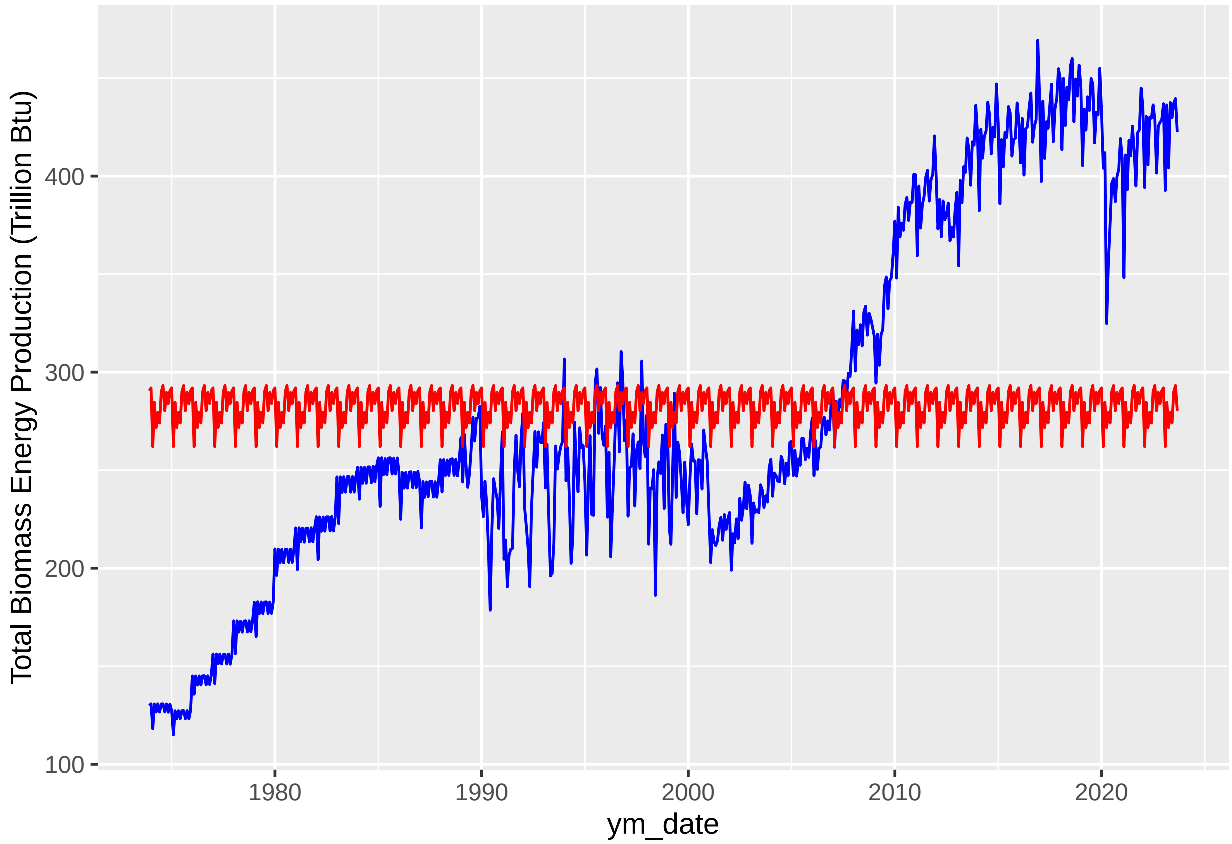
```
##
## Call:
## lm(formula = ts_energy_data[, 3] ~ hydro_dummies)
```

```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -31.500  -5.842  -0.512   6.410  32.314
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)        69.404      1.496  46.389  < 2e-16 ***
## hydro_dummiesJan   10.878      2.105   5.167 3.26e-07 ***
## hydro_dummiesFeb   15.596      2.105   7.408 4.49e-13 ***
## hydro_dummiesMar    8.113      2.105   3.854 0.000129 ***
## hydro_dummiesApr   17.679      2.105   8.398 3.44e-16 ***
## hydro_dummiesMay   16.246      2.105   7.717 5.16e-14 ***
## hydro_dummiesJun   24.971      2.105  11.862  < 2e-16 ***
## hydro_dummiesJul   21.705      2.105  10.310  < 2e-16 ***
## hydro_dummiesAug   14.926      2.105   7.090 3.87e-12 ***
## hydro_dummiesSep    5.234      2.105   2.486 0.013182 *
## hydro_dummiesOct   -5.830      2.105  -2.769 0.005793 **
## hydro_dummiesNov   -5.573      2.116  -2.634 0.008666 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.47 on 586 degrees of freedom
## Multiple R-squared:  0.467,  Adjusted R-squared:  0.457
## F-statistic: 46.68 on 11 and 586 DF,  p-value: < 2.2e-16
```

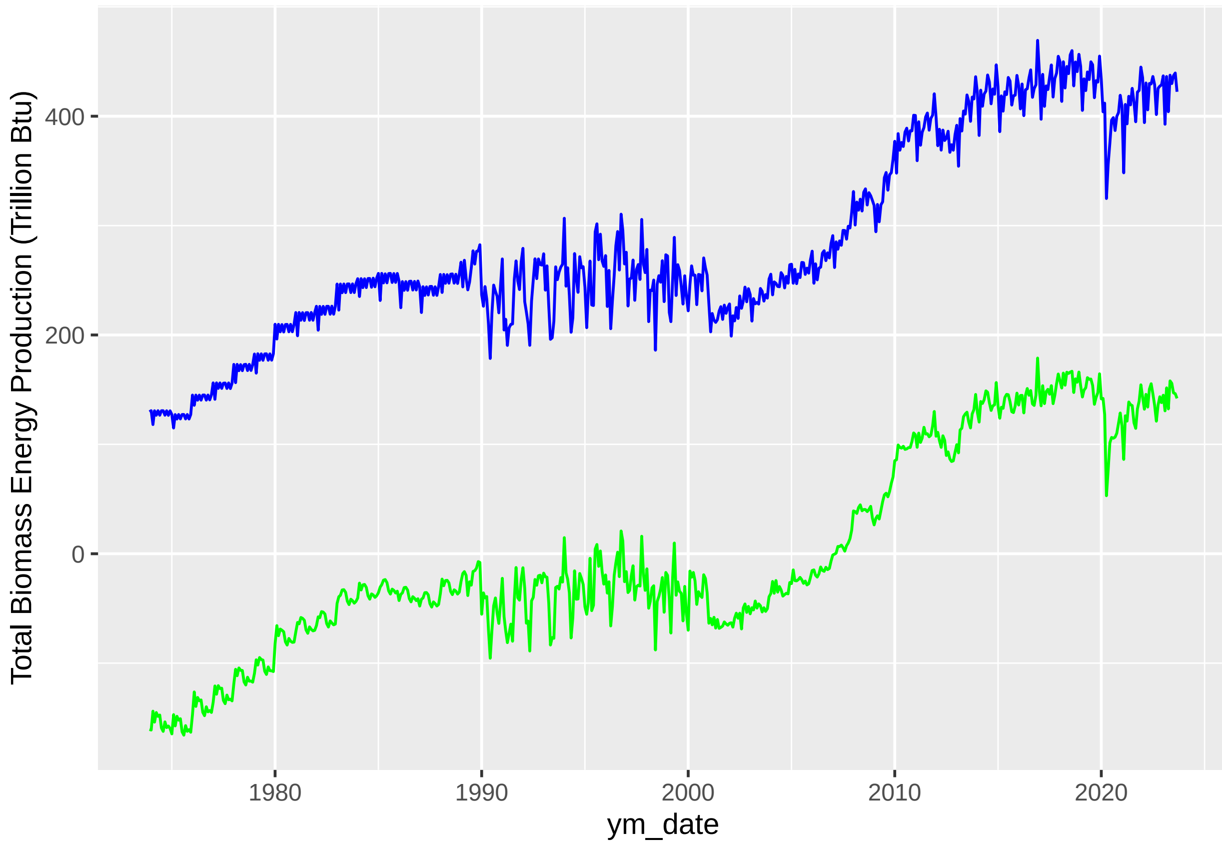
```
#Store regression coefficients
hydro_beta_int <- hydro_seas_means_model$coefficients[1]
hydro_beta_coeff <- hydro_seas_means_model$coefficients[2:12]
```

The coefficients indicate a very strong seasonalcomponent to the data, with increased values in the summer and decreased values in the fall and winter months.These coefficients have a strong R square value and all have strong p values as well. This series has the most significant seasonal component of the three. ### 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?

```
#compute seasonal component
biomass_seas_comp <- array(0,n_obs)
for(i in 1:n_obs){
  biomass_seas_comp[i] <- (biomass_beta_int+biomass_beta_coeff
                           %*% biomass_dummies[i,])
}
#Viewing seasonal and original
ggplot(energy_data, aes(x=ym_date, y=energy_data[,1])) +
            geom_line(color="blue") +
            ylab(paste0(colnames(energy_data)[1])) +
            geom_line(aes(y=biomass_seas_comp), col="red")
```

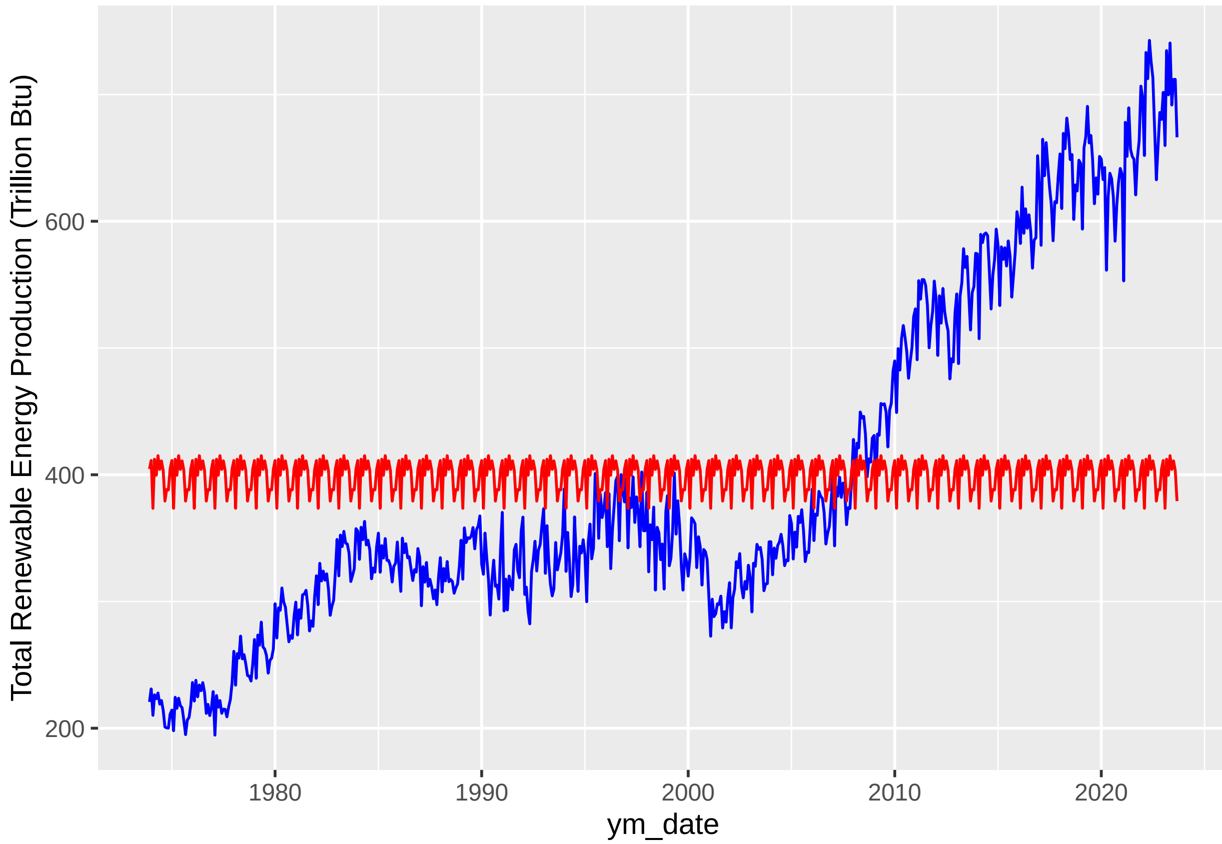



```r
#Removing seasonal component
deseason_biomass_data <- energy_data[,1]-biomass_seas_comp

#Plotted together
ggplot(energy_data, aes(x=ym_date, y=energy_data[,1])) +
            geom_line(color="blue") +
            ylab(paste0(colnames(energy_data)[1])) +
            geom_line(aes(y=deseason_biomass_data), col="green")
```

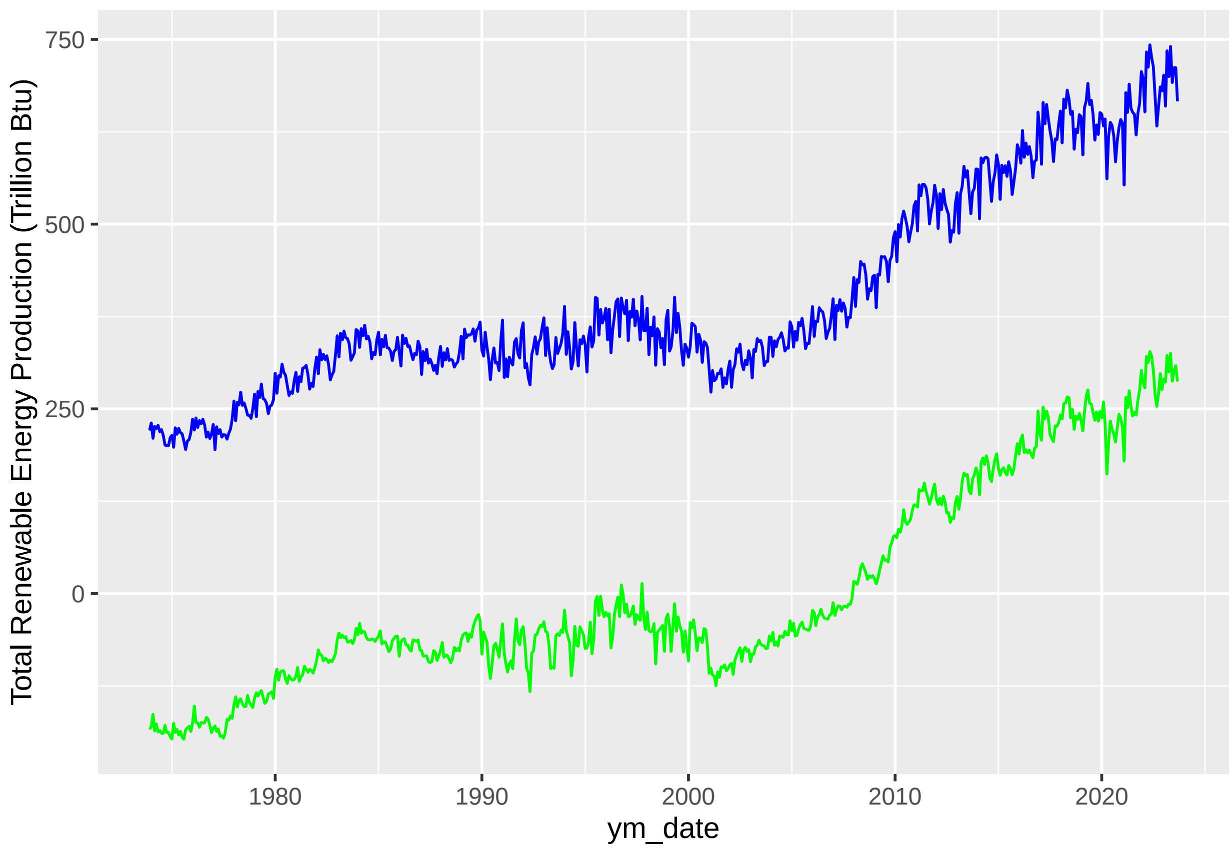


Biomass appears to have a very small seasonal trend that was removed through this process that smoothed out the regular variations year to year. De-seasoning the data also centered the mean at zero.

```
#compute seasonal component
RE_seas_comp <- array(0,n_obs)
for(i in 1:n_obs){
  RE_seas_comp[i] <- (RE_beta_int+RE_beta_coeff %*% RE_dummies[i,])
}
#Viewing seasonal and original
ggplot(energy_data, aes(x=ym_date, y=energy_data[,2])) +
            geom_line(color="blue") +
            ylab(paste0(colnames(energy_data)[2])) +
            geom_line(aes(y=RE_seas_comp), col="red")
```

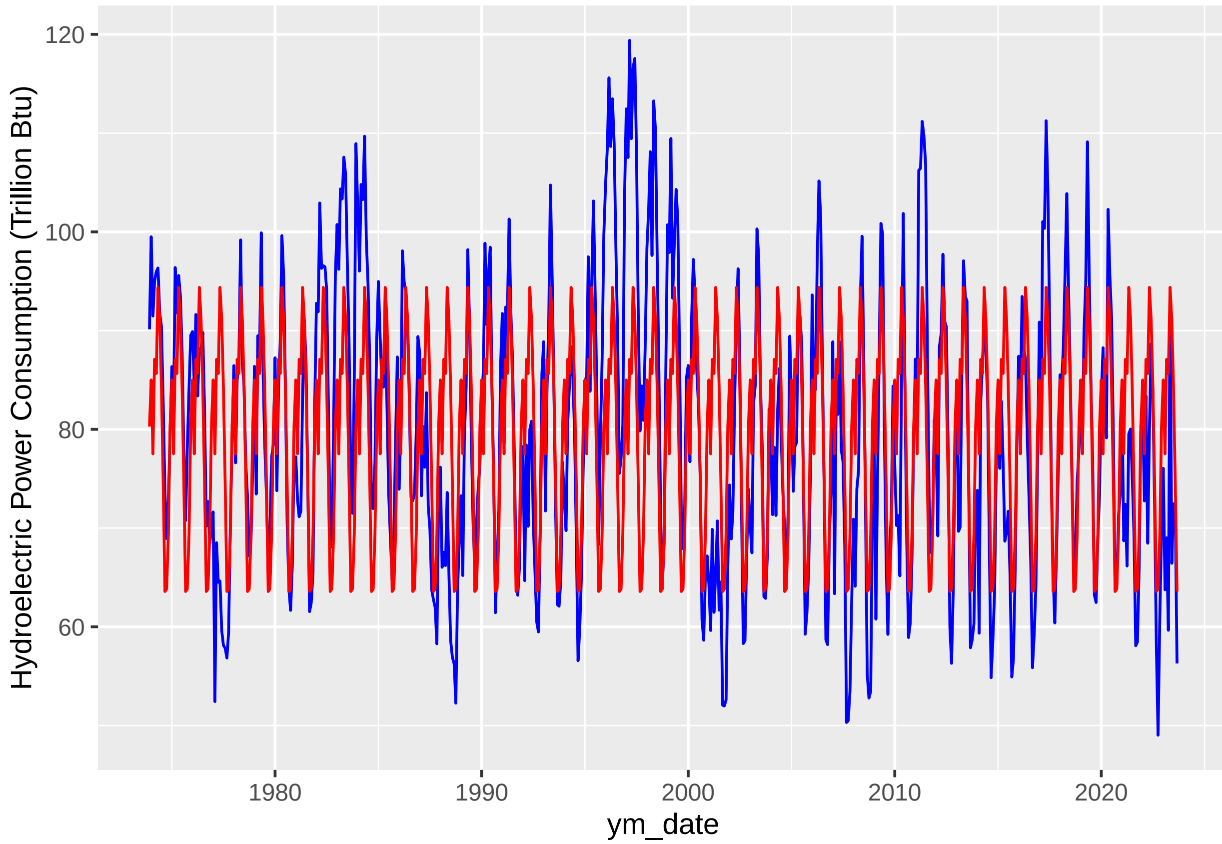



```
#Removing seasonal component
deseason_RE_data <- energy_data[,2]-RE_seas_comp

#Plotted together
ggplot(energy_data, aes(x=ym_date, y=energy_data[,2])) +
            geom_line(color="blue") +
            ylab(paste0(colnames(energy_data)[2])) +
            geom_line(aes(y=deseason_RE_data), col="green")
```

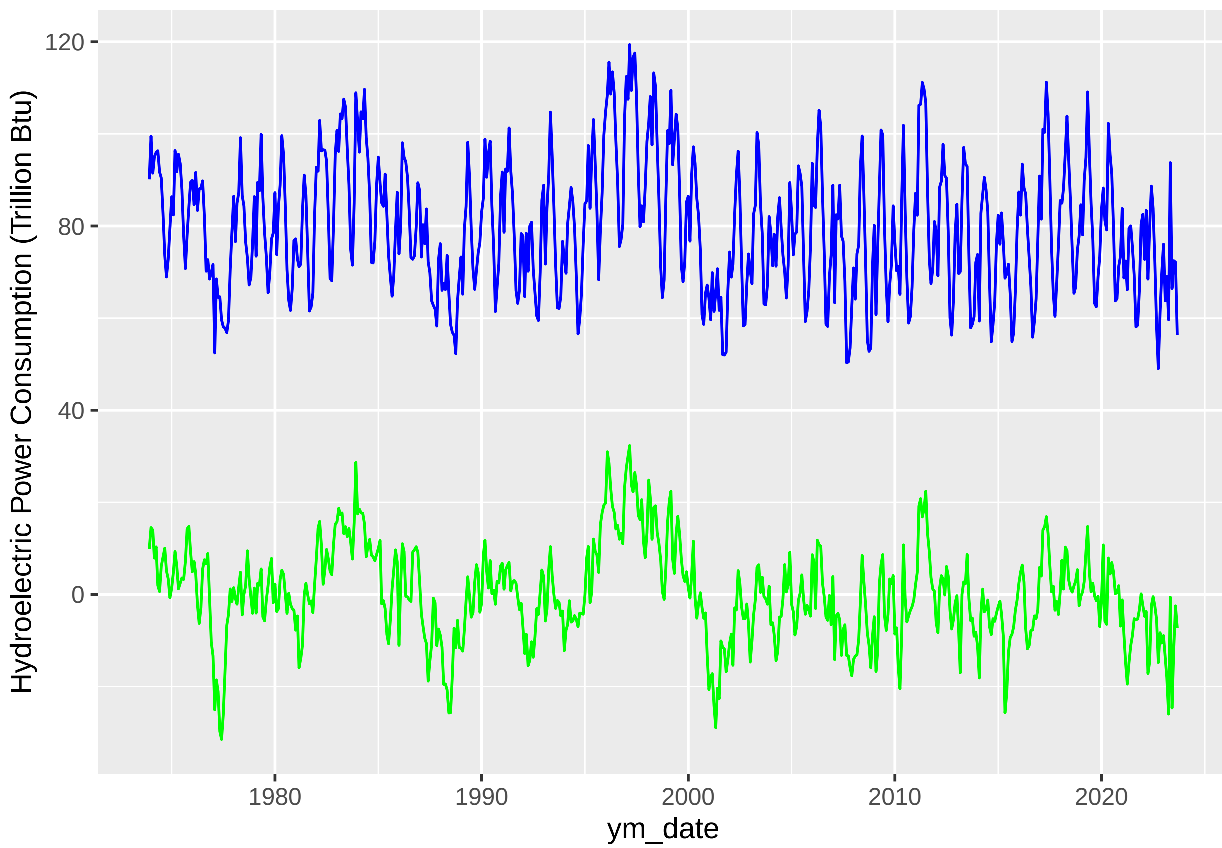


RE appears to have a very small seasonal trend that was removed through this process that smoothed out the regular variations year to year. De-seasoning the data also centered the mean at zero.

```
#compute seasonal component
hydro_seas_comp <- array(0,n_obs)
for(i in 1:n_obs){
  hydro_seas_comp[i] <- (hydro_beta_int+hydro_beta_coeff %*% hydro_dummies[i,])
}
#Viewing seasonal and original
ggplot(energy_data, aes(x=ym_date, y=energy_data[,3])) +
            geom_line(color="blue") +
            ylab(paste0(colnames(energy_data)[3])) +
            geom_line(aes(y=hydro_seas_comp), col="red")
```




```r
#Removing seasonal component
deseason_hydro_data <- energy_data[,3]-hydro_seas_comp

#Plotted together
ggplot(energy_data, aes(x=ym_date, y=energy_data[,3])) +
            geom_line(color="blue") +
            ylab(paste0(colnames(energy_data)[3])) +
            geom_line(aes(y=deseason_hydro_data), col="green")
```



Deseasoning the hydro series smoothed out the regular yearly peaks and vallys in the hydro data. It also centered the mean at 0. This plot had the greatest change.

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?

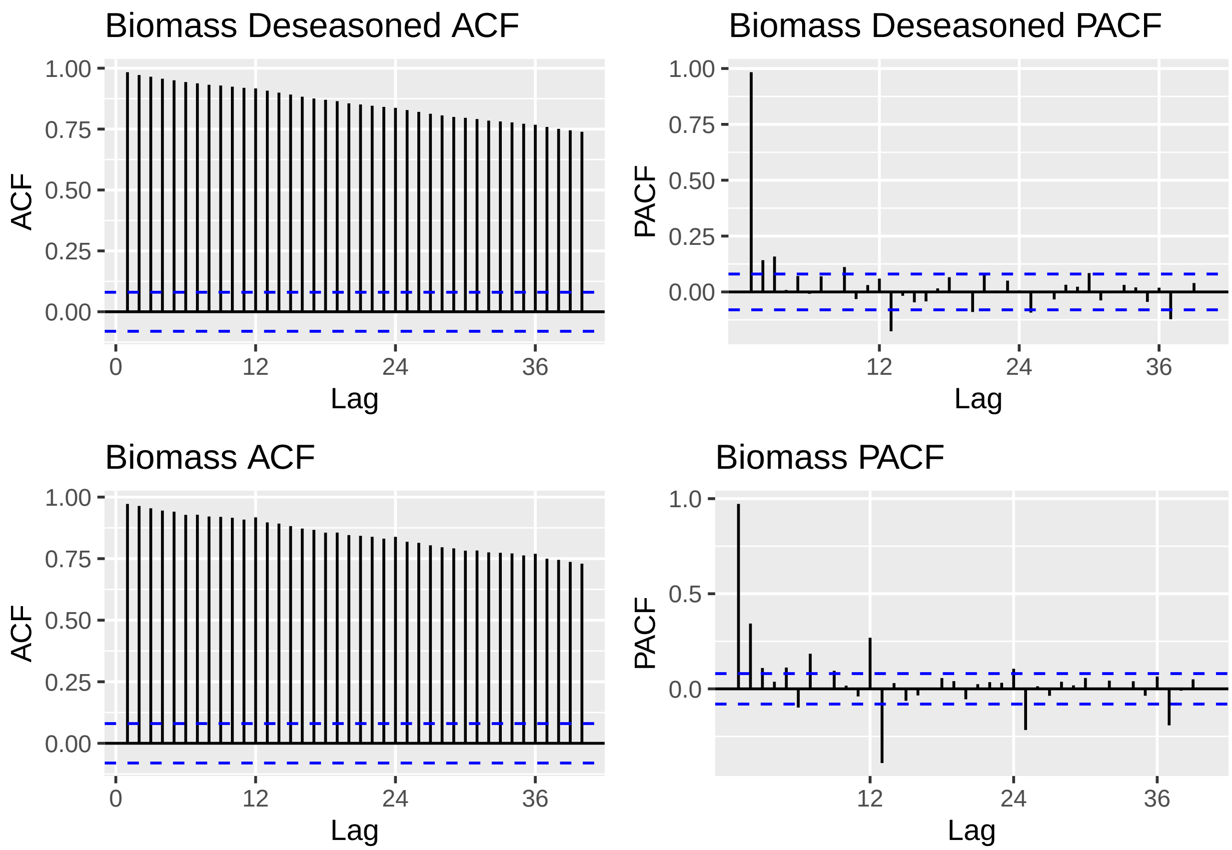
```r
ts_biomass_deseason <- ts(deseason_biomass_data,start=c(1973,1),frequency = 12)
ts_RE_deseason <- ts(deseason_RE_data,start=c(1973,1),frequency = 12)
ts_hydro_deseason <- ts(deseason_hydro_data,start=c(1973,1),frequency = 12)
```

```r
biomass_deseason_acf <- ggAcf(ts_biomass_deseason,lag.max=40)+
  labs(title="Biomass Deseasoned ACF")
#biomass_season_acf
hydro_deseason_acf <- ggAcf(ts_hydro_deseason,lag.max=40)+
  labs(title="Hydro Deseasoned ACF")
#hydro_season_acf
RE_deseason_acf <- ggAcf(ts_RE_deseason,lag.max=40)+
  labs(title="RE Deseasoned ACF")
#RE_season_acf
```

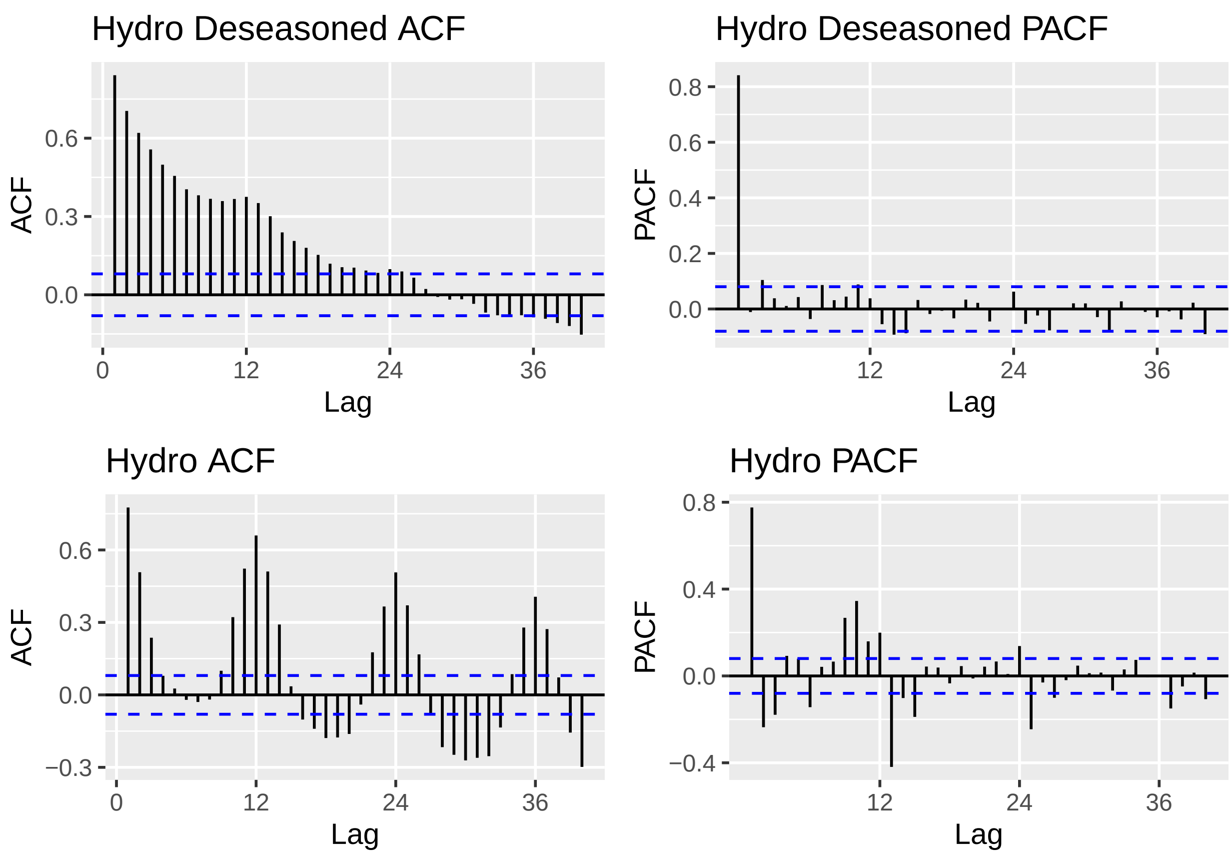
```r
biomass_deseason_pacf <- ggPacf(ts_biomass_deseason,lag.max=40)+
  labs(title="Biomass Deseasoned PACF")
#biomass_season_pacf
```

```
hydro_deseason_pacf <- ggPacf(ts_hydro_deseason,lag.max=40)+
  labs(title="Hydro Deseasoned PACF")
#hydro_season_pacf
RE_deseason_pacf <- ggPacf(ts_RE_deseason,lag.max=40)+
  labs(title="RE Deseasoned PACF")
#RE_season_pacf
```

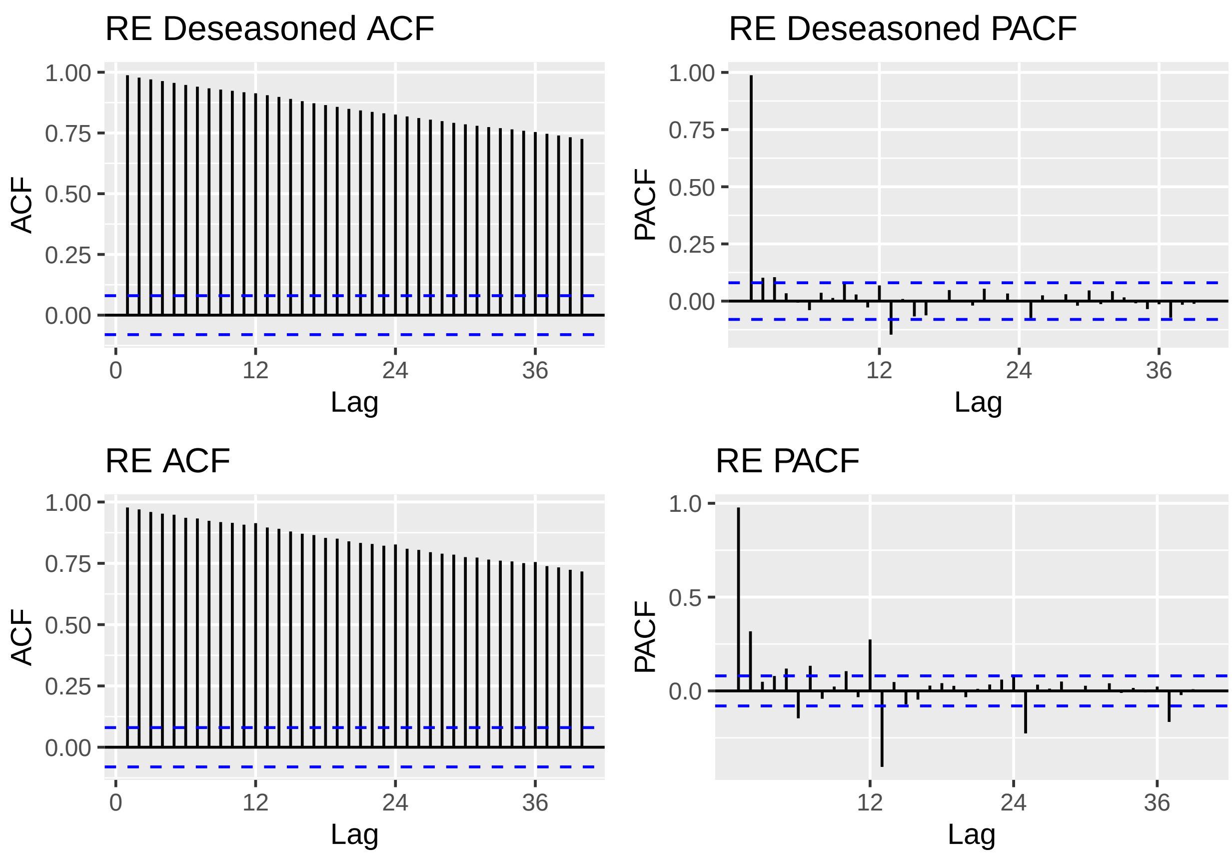
```
plot_grid(biomass_deseason_acf,biomass_deseason_pacf,biomass_acf,biomass_pacf,
          ncol=2,align="h")
```




```
plot_grid(hydro_deseason_acf,hydro_deseason_pacf ,hydro_acf,hydro_pacf,
          ncol=2,align="h")
```




```
plot_grid(RE_deseason_acf,RE_deseason_pacf,RE_acf,RE_pacf,
          ncol=2,align="h")
```



The biomass and renewable energy PACF and ACFs look very similar to their deseasoned versions. However, the hydro deseasoned ACF shows far less correlation to previous data due to taking out the seasonal similarities,and the PACF has an even smaller statistical significance in probability the further along the lag goes. Taking out the seasonal component really helped the hydro data to be less dependent on many periods before it.