ENV 790.30 - Time Series Analysis for Energy Data | Spring 2022

Assignment 3 - Due date 02/08/22

Sarah Mansfield

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 project open the first thing you will do is change "Student Name" on line 3 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. Rename the pdf file such that it includes your first and last name (e.g., "LuanaLima_TSA_A03_Sp22.Rmd"). 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 January 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)
library(tseries)
library(Kendall)
library(dplyr)
library(readxl)
library(ggfortify)
library(patchwork)
```

```
#Importing data set
energy <- read_excel("~/ENVIRON 790/ENV790_TimeSeriesAnalysis_Sp2022/Data/Table_10.1_Renewable_Energy_P
                     skip = 10)
```

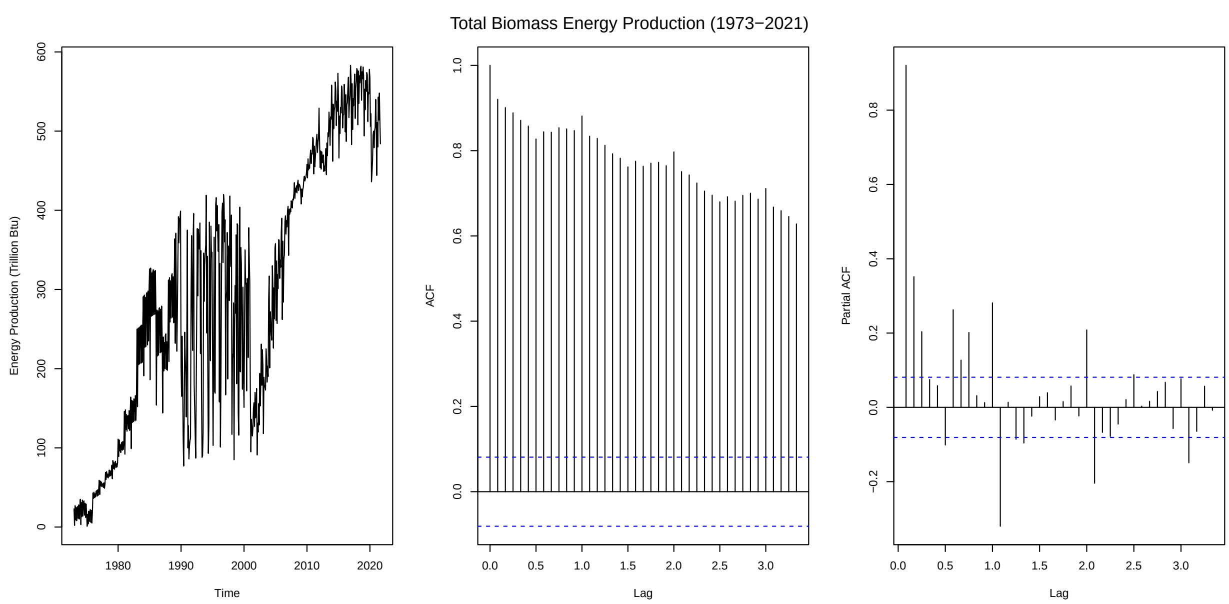
```
energy <- energy %>%
  slice(2:nrow(energy)) %>% # delete extraneous first row
  select(`Total Biomass Energy Production`,
         `Total Renewable Energy Production`,
         `Hydroelectric Power Consumption`)
energy_ts <- ts(energy, start = c(1973, 1), frequency = 12)
```

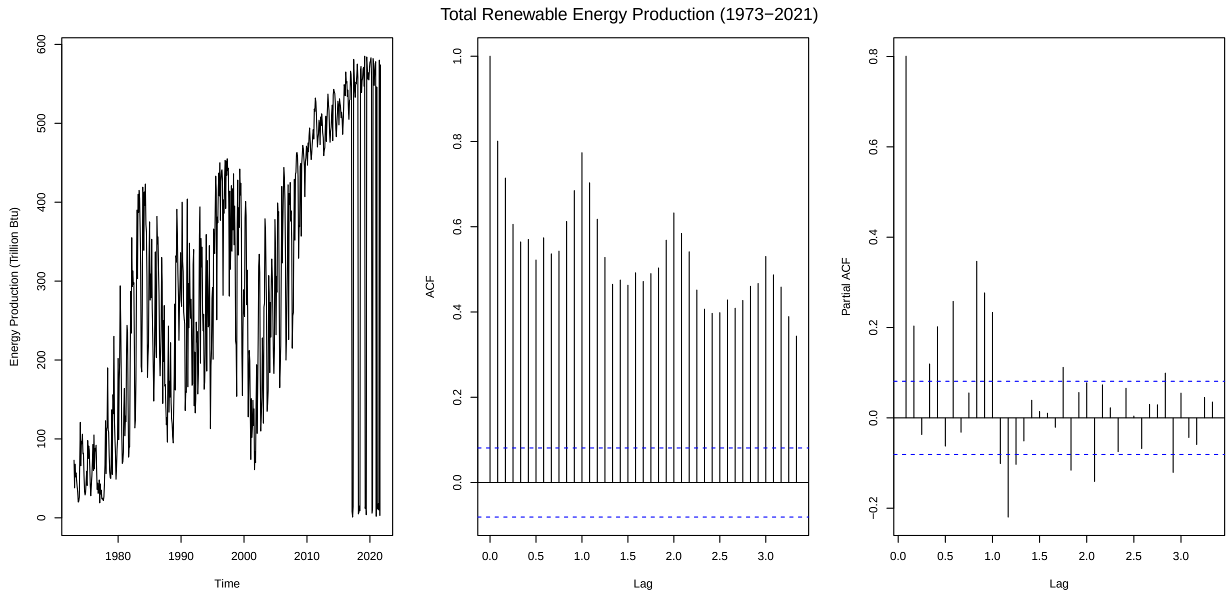
Trend Component

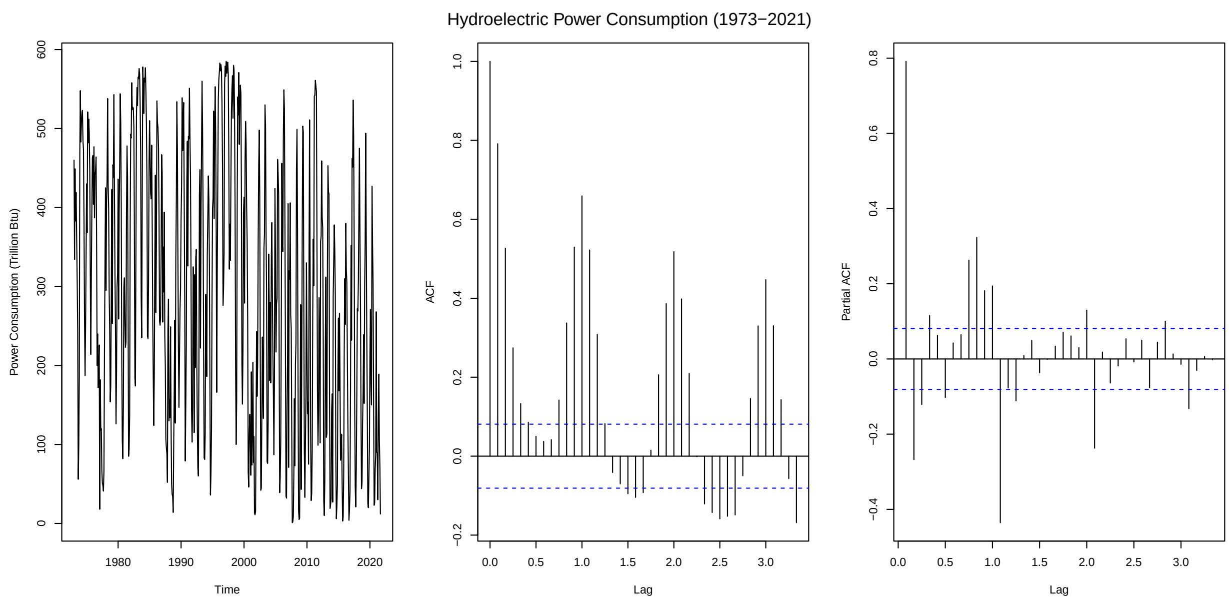
Q1

Create a plot window that has one row and three columns. And then for each object on your data frame, fill the plot window with time series plot, ACF and PACF. You may use the some code form A2, but I want all three plots on the same window this time. (Hint: use par() function)

```
series <- colnames(energy_ts)
for (i in 1:ncol(energy_ts)) {
  par(mfrow=c(1,3))
  ylabel <- ifelse(series[i] == "Hydroelectric Power Consumption",
                   "Power Consumption (Trillion Btu)",
                   "Energy Production (Trillion Btu)")
  plot(energy_ts[,i], main = "",
     xlab = "Time", ylab = ylabel)
  acf(energy_ts[,i], lag.max = 40, main = "")
  pacf(energy_ts[,i], lag.max = 40, main = "")
  mtext(paste0(series[i], " (1973-2021)"), side = 3, line = -3, outer = TRUE)
}
```







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?

Yes - both Total Biomass Energy Production and Total Renewable Energy Production have an overall increasing trend over time, whereas Hydroelectric Power Consumption has an overall decreasing trend (although note that the magnitude of this trend is much less apparent compared to the other two).

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.

```r
t <- c(1:nrow(energy_ts))

coefs <- list()
for (i in 1:length(series)) {
  lm_mod <- lm(energy_ts[,i] ~ t)
  coefs[[i]] <- list(beta0 = as.numeric(lm_mod$coefficients[1]),
                     beta1 = as.numeric(lm_mod$coefficients[2]))
  print(series[i])
  print(summary(lm_mod))
}
```

```
## [1] "Total Biomass Energy Production"
##
## Call:
## lm(formula = energy_ts[, i] ~ t)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -250.02  -35.45    8.29   52.77  184.62
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 37.44854    6.88821   5.437    8e-08 ***
## t            0.86734    0.02037  42.583   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 83.2 on 583 degrees of freedom
## Multiple R-squared:  0.7567, Adjusted R-squared:  0.7563
## F-statistic:  1813 on 1 and 583 DF,  p-value: < 2.2e-16
##
## [1] "Total Renewable Energy Production"
##
## Call:
## lm(formula = energy_ts[, i] ~ t)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -483.59  -68.23   25.07   88.65  233.78
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 98.07627   10.45701   9.379   <2e-16 ***
## t            0.66527    0.03092  21.515   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 126.3 on 583 degrees of freedom
## Multiple R-squared:  0.4426, Adjusted R-squared:  0.4416
## F-statistic: 462.9 on 1 and 583 DF,  p-value: < 2.2e-16
##
## [1] "Hydroelectric Power Consumption"
```

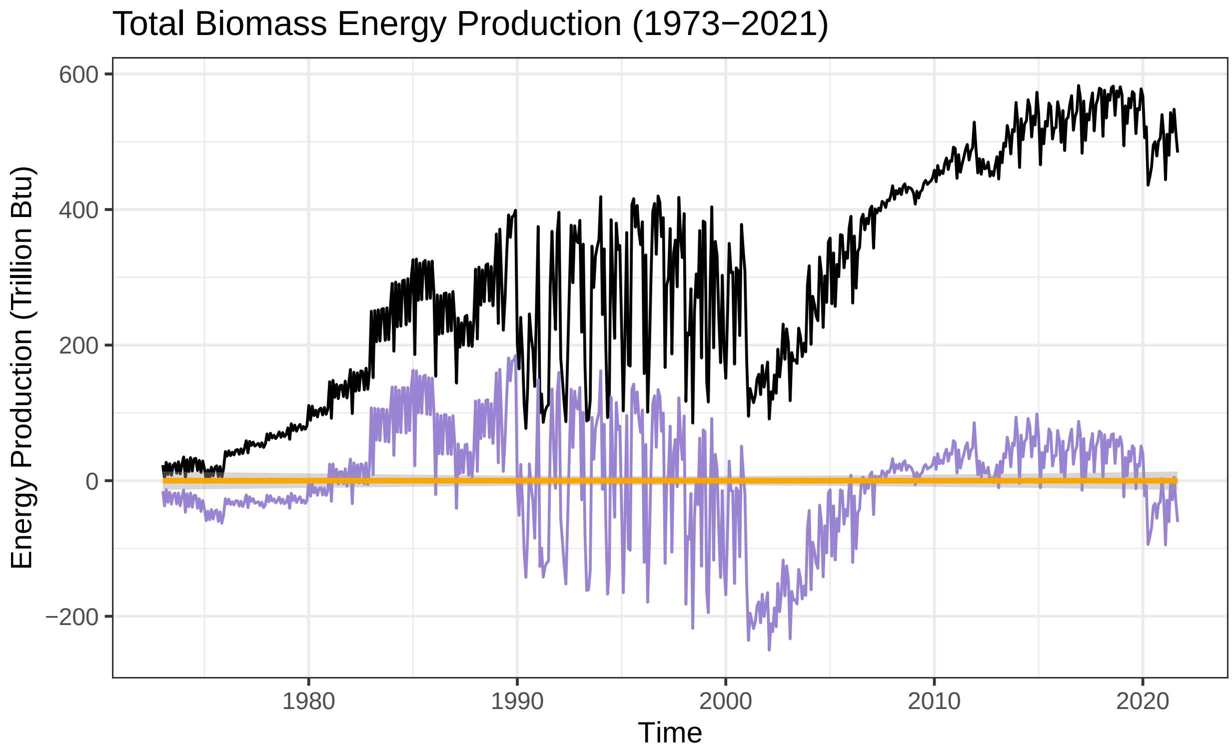
```
##
## Call:
## lm(formula = energy_ts[, i] ~ t)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -351.60 -135.20    2.79  125.90  320.96
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 385.3595    13.2920   28.99  < 2e-16 ***
## t            -0.3152     0.0393   -8.02 5.82e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 160.5 on 583 degrees of freedom
## Multiple R-squared:  0.09936,    Adjusted R-squared:  0.09782
## F-statistic: 64.32 on 1 and 583 DF,  p-value: 5.821e-15
```

- When $t = 0$ (the month directly preceding Jan 1973), the expected Total Biomass Energy Production is approximately 37.44854 Trillion Btu. For each one unit increase in time (months), the estimated Total Biomass Energy Production is expected to increase by approximately 0.86734 Trillion Btu.
- When $t = 0$ (the month directly preceding Jan 1973), the expected Total Renewable Energy Production is approximately 98.07627 Trillion Btu. For each one unit increase in time (months), the estimated Total Renewable Energy Production is expected to increase by approximately 0.66527 Trillion Btu.
- When $t = 0$ (the month directly preceding Jan 1973), the expected Hydroelectric Power Consumption is approximately 385.3595 Trillion Btu. For each one unit increase in time (months), the estimated Hydroelectric Power Consumption is expected to decrease by approximately 0.3152 Trillion Btu.

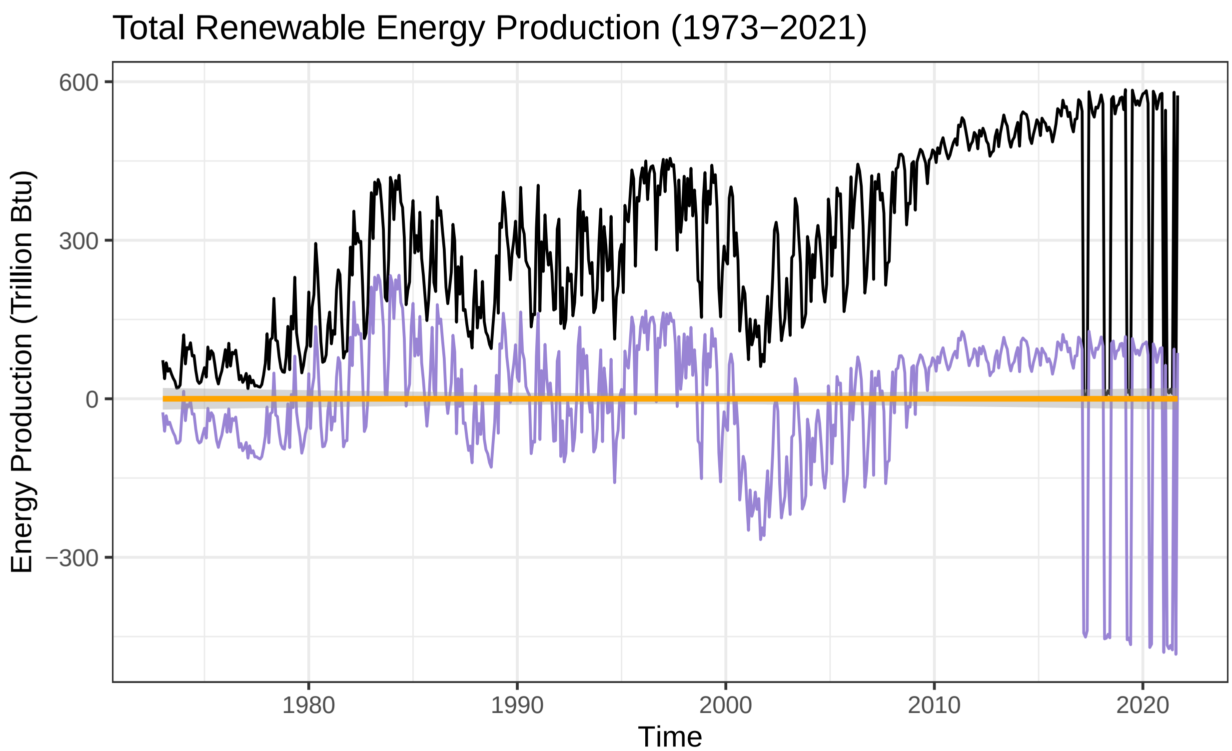
Q4

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?

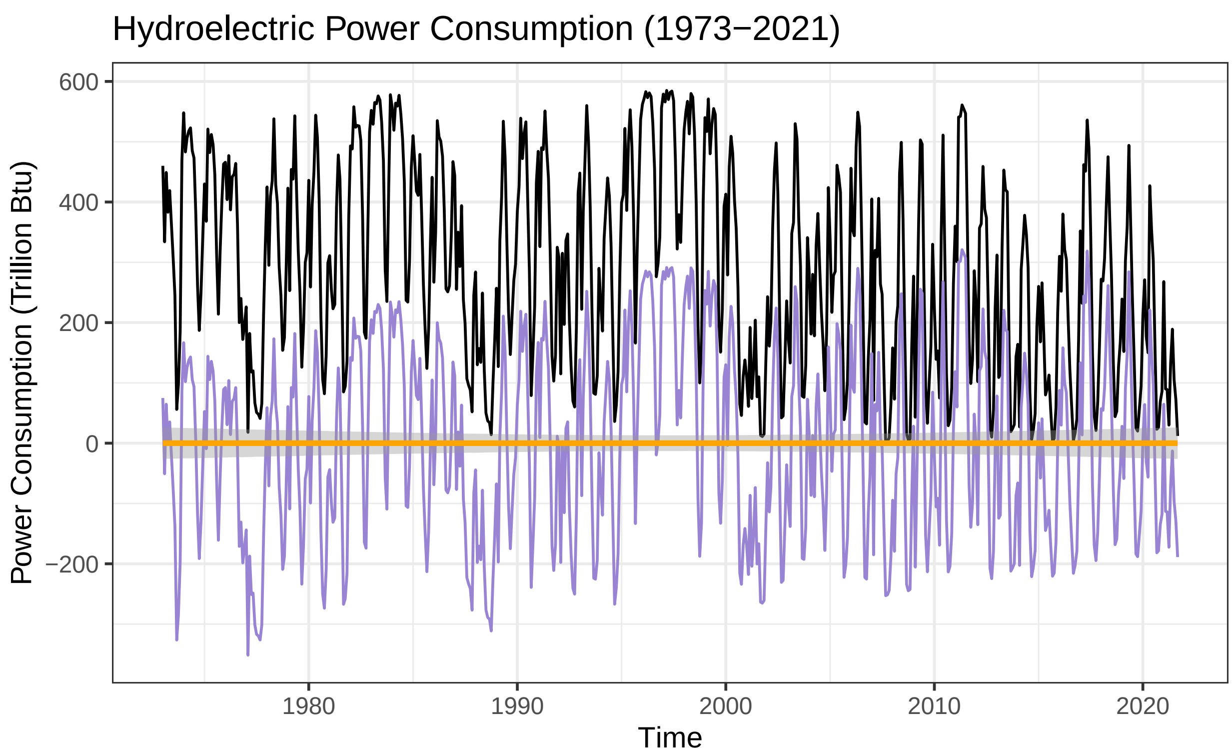
```r
for (i in 1:ncol(energy_ts)) {
  coef_series <- coefs[[i]]
  detrend_series <- energy_ts[,i] - (coef_series$beta0 + coef_series$beta1 * t)
  ylabel <- ifelse(series[i] == "Hydroelectric Power Consumption",
                   "Power Consumption (Trillion Btu)",
                   "Energy Production (Trillion Btu)")
  print(
    autoplot(energy_ts[,i]) +
      geom_line(aes(y = detrend_series), col = "#9984d4") +
      geom_smooth(aes(y = detrend_series), color = "orange", method = "lm") +
      labs(title = paste0(series[i], " (1973-2021)"),
           x = "Time", y = ylabel) +
      theme_bw()
    )
}
```



Total Biomass Energy Production (1973−2021)
Energy Production (Trillion Btu)
600
400
200
0
−200
1980
1990
2000
2010
2020
Time



Total Renewable Energy Production (1973−2021)
Energy Production (Trillion Btu)
600
300
0
−300
1980
1990
2000
2010
2020
Time



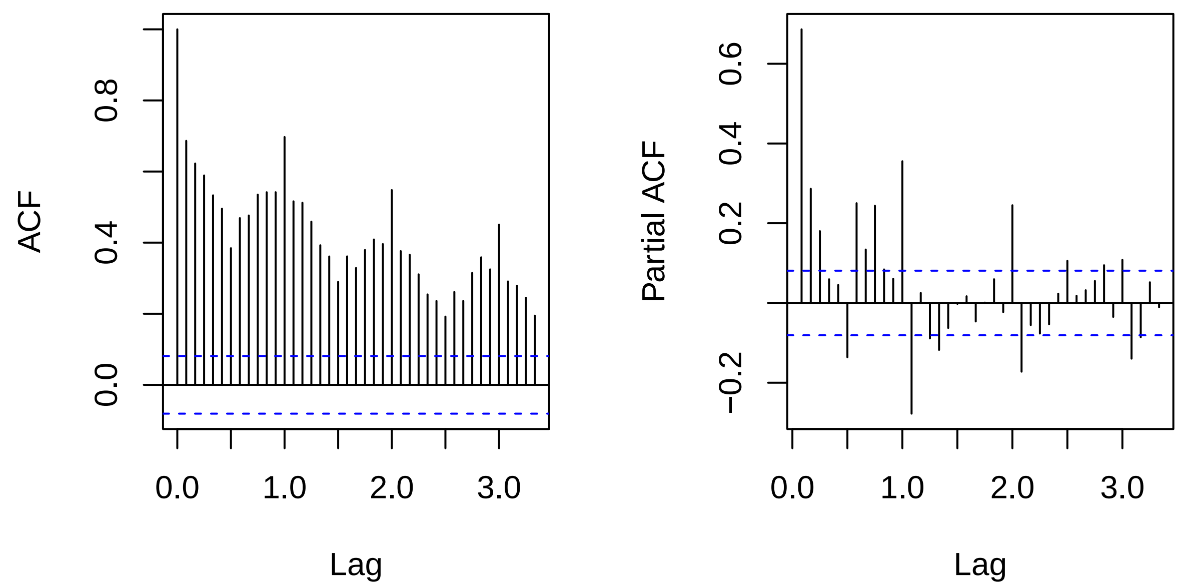
For all three detrended series, the plots show each having a slope of 0, indicating that we were able to effectively eliminate the trend with a linear model. For Total Biomass Energy Production and Total Renewable Energy Production, the two series that had an overall increasing trend, the difference between the original and the detrended series widens as time increases. For Hydroelectric Power Consumption we see the opposite effect, with the difference lessening as time increases.

Q5

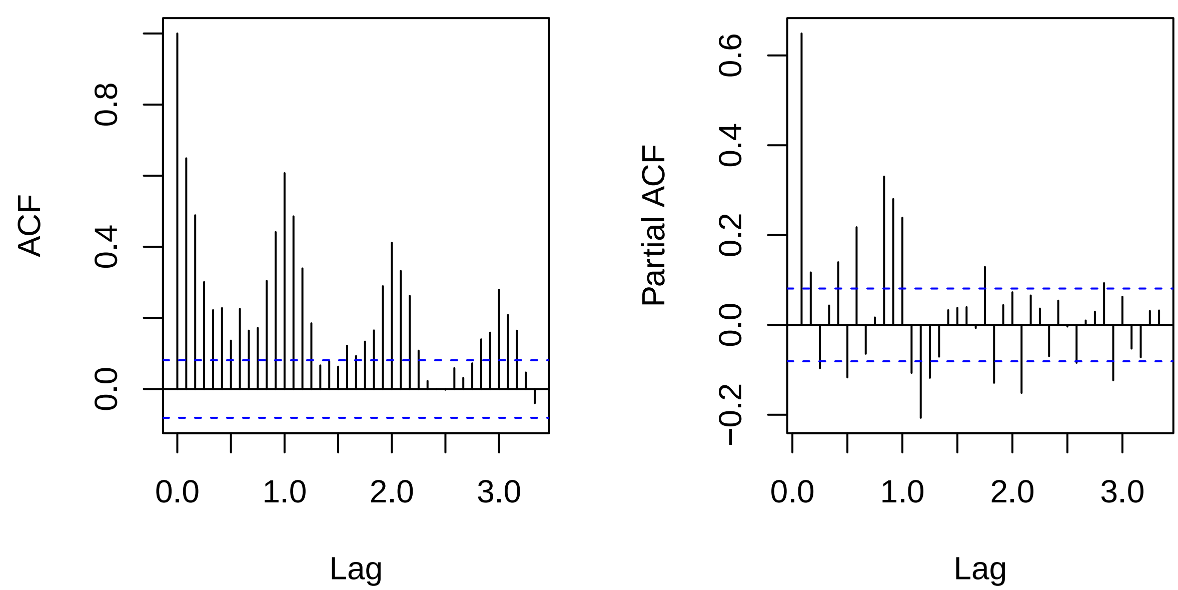
Plot ACF and PACF for the detrended series and compare with the plots from Q1. Did the plots change? How?

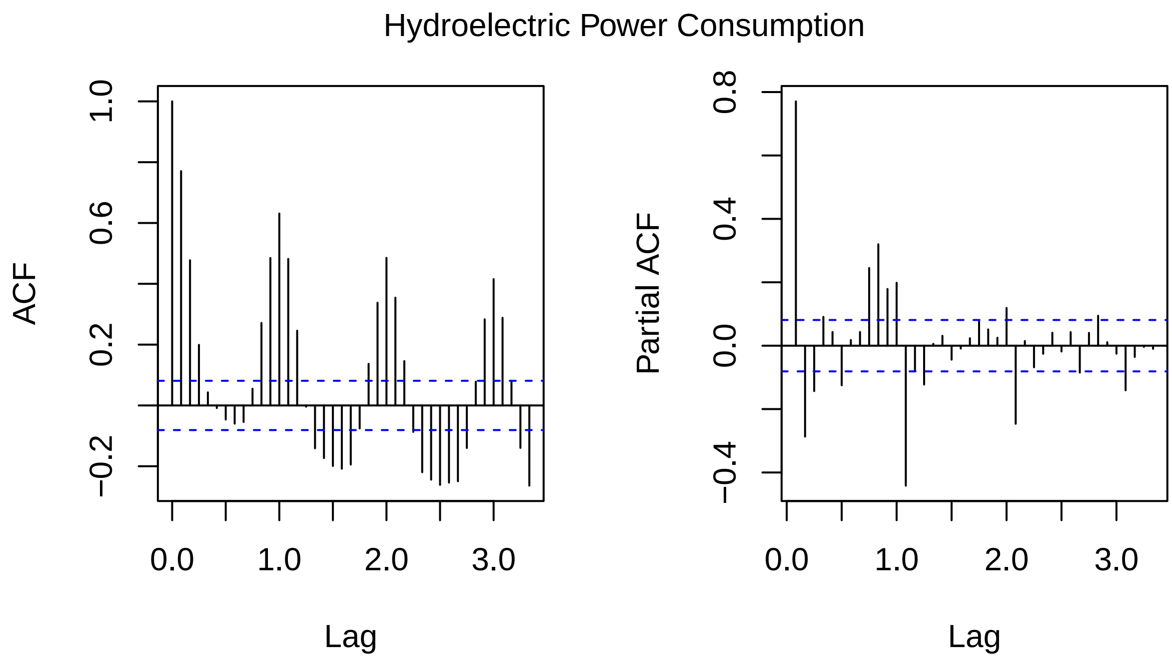
```
for (i in 1:ncol(energy_ts)) {
  par(mfrow=c(1,2))
  coef_series <- coefs[[i]]
  detrend_series <- energy_ts[,i] - (coef_series$beta0 + coef_series$beta1 * t)
  acf(detrend_series, lag.max = 40, main = "")
  pacf(detrend_series, lag.max = 40, main = "")
  mtext(series[i], side = 3, line = -3, outer = TRUE)
}
```

Total Biomass Energy Production



Total Renewable Energy Production





Compared with the plots from Q1, the PACF plots for all three series look very similar. On the other hand, comparing the ACF plots, we see that for Total Biomass Energy Production and Total Renewable Energy Production the ACF values drop at a sharper and quicker rate compared to the detrended series. This makes sense because these two series exhibited a steeper trend as compared to Hydroelectric Power Consumption, and therefore by removing the trend effect we also removed more of the autocorrelation between units of time.

Seasonal Component

Set aside the detrended series and consider the original series again from Q1 to answer Q6 to Q8.

Q6

Do the series seem to have a seasonal trend? Which serie/series? Use function *lm()* to fit a seasonal means model (i.e. using the seasonal dummies) to this/these time series. Ask R to print the summary of the regression. Interpret the regression output. Save the regression coefficients for further analysis.

Hydroelectric Power Consumption seems to be the only one out of the three series that exhibits a seasonal trend. We fit a seasonal means model to this time series as follows:

```
dummies <- seasonaldummy(energy_ts[,3])
seas_means_model <- lm(energy_ts[,3] ~ dummies)
summary(seas_means_model)
```

```
##
## Call:
## lm(formula = energy_ts[, 3] ~ dummies)
##
```

```
## Residuals:
##      Min       1Q  Median       3Q      Max
## -368.22  -86.33   -6.92   89.98  304.49
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   303.92      18.57  16.366  < 2e-16 ***
## dummiesJan     55.31      26.13   2.117  0.03470 *
## dummiesFeb    -36.41      26.13  -1.393  0.16403
## dummiesMar     68.53      26.13   2.623  0.00895 **
## dummiesApr     64.88      26.13   2.483  0.01331 *
## dummiesMay    144.31      26.13   5.523 5.05e-08 ***
## dummiesJun    117.10      26.13   4.482 8.93e-06 ***
## dummiesJul     46.16      26.13   1.767  0.07777 .
## dummiesAug    -72.86      26.13  -2.788  0.00547 **
## dummiesSep   -198.59      26.13  -7.601 1.21e-13 ***
## dummiesOct   -192.77      26.26  -7.340 7.34e-13 ***
## dummiesNov   -132.65      26.26  -5.051 5.92e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 128.7 on 573 degrees of freedom
## Multiple R-squared:  0.4315, Adjusted R-squared:  0.4206
## F-statistic: 39.54 on 11 and 573 DF,  p-value: < 2.2e-16
```

```
beta_int <- seas_means_model$coefficients[1]
beta_coeff <- seas_means_model$coefficients[2:12]
```

For the month of December, the expected Hydroelectric Power Consumption is approximately 303.92 Trillion Btu. In January, this expected value increases by approximately 55.31 to equal 359.23 Trillion Btu. In February the expected value is approximately $303.92 - 36.41 = 267.51$ Trillion Btu. For each month we can thus calculate the expected Hydroelectric Power Consumption by adding the intercept estimate and the relevant dummy variable estimate.

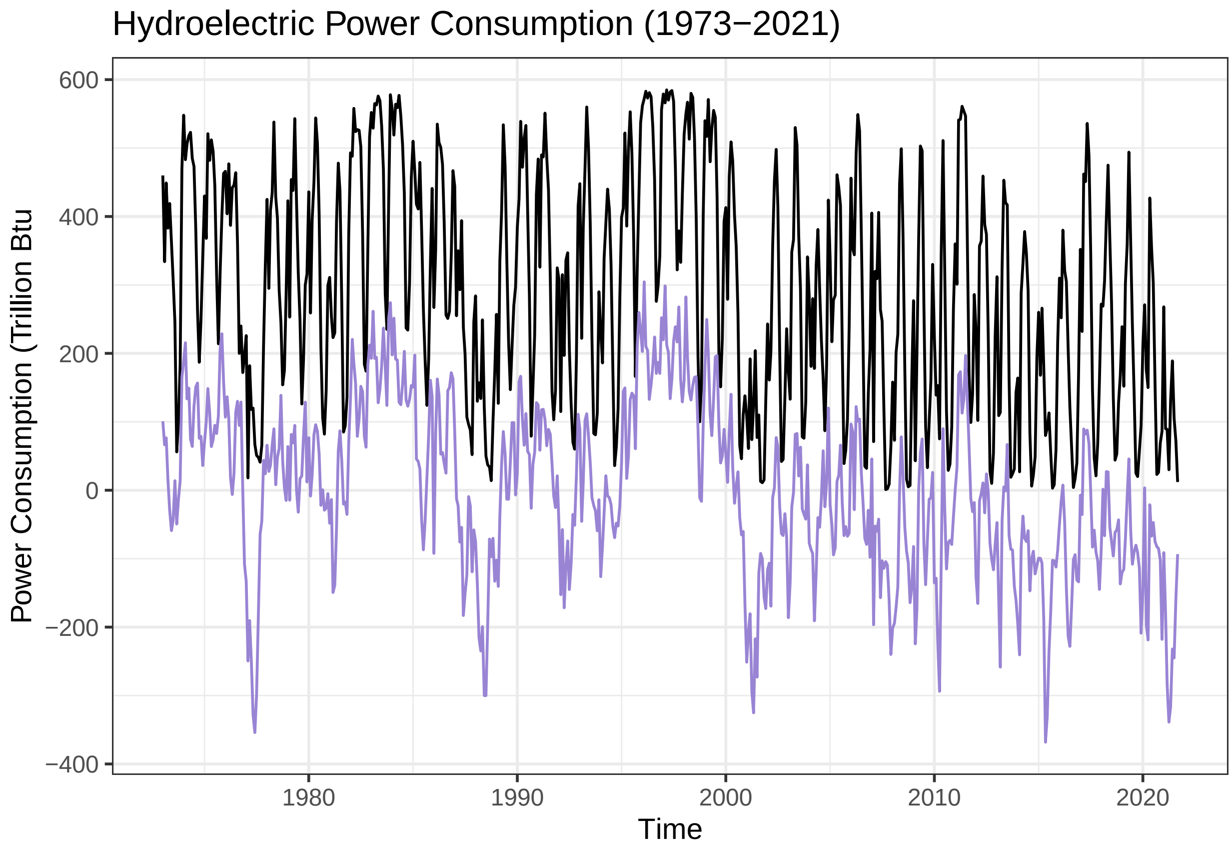
Q7

Use the regression coefficients from Q6 to deseason the series. Plot the deseason series and compare with the plots from part Q1. Did anything change?

```
# compute seasonal component
inflow_seas_comp <- array(0, nrow(energy_ts))
for (i in 1:nrow(energy_ts)) {
  inflow_seas_comp[i] <- (beta_int + beta_coeff %*% dummies[i,])
}

# removing seasonal component
deseason_energy <- energy_ts[,3] - inflow_seas_comp

autoplot(energy_ts[,3]) +
  geom_line(aes(y = deseason_energy), col = "#9984d4") +
  labs(title = paste0(series[3], " (1973-2021)"),
       x = "Time", y = "Power Consumption (Trillion Btu") +
  theme_bw()
```

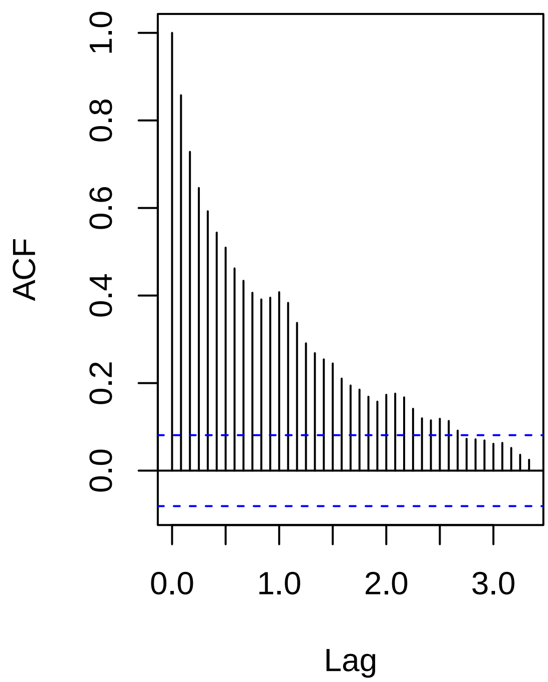


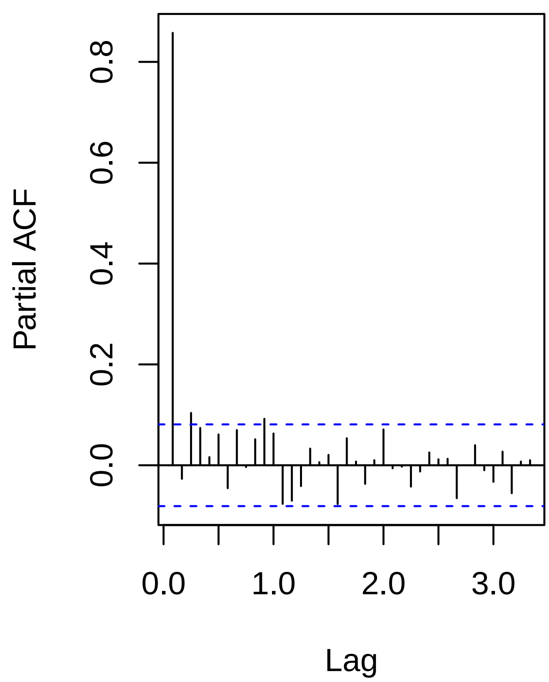
After removing the seasonal component and comparing the deseason series to the original series, we can see that the deseason series looks to have a less regular trend over time, with there being more variation in dips and peaks in the data. Additionally, it looks like the deseason series has an overall negative trend that is more prominent when compared to the original series.

Q8

Plot ACF and PACF for the deseason series and compare with the plots from Q1. Did the plots change? How?

```
par(mfrow=c(1,2))
acf(deseason_energy, lag.max = 40, main = "")
pacf(deseason_energy, lag.max = 40, main = "")
```





For the ACF plot, the deseason series exhibits a much more gradual decrease in ACF values, unlike the original series that had a sharp initial drop as well as a pattern of increasing and decreasing ACF values (the bell curve like shape as mentioned in Q1). Comparing the PACF plots, the deseason series has a sharp initial drop much like the original, but after the initial drop its PACF values remain small with little fluctuation as lag increases (the original series had a greater fluctuation in PACF values as the lag increased).