ENV 797 - Time Series Analysis for Energy and Environment Applications | Spring 2026

Assignment 4 - Due date 02/10/26

Joy Wu

Directions

You should open the .rmd file corresponding to this assignment on RStudio. The file is available on our class repository on Github. And to do so you will need to fork our repository and link it to your RStudio.

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_A04_Sp26.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).

When you have completed the assignment, **Knit** the text and code into a single PDF file. Submit this pdf using Sakai.

R packages needed for this assignment: "xlsx" or "readxl", "ggplot2", "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(ggplot2)
library(forecast)
library(tseries)
library(Kendall)
library(cowplot)
```

Questions

Consider the same data you used for A3 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. **For this assignment you will work only with the column "Total Renewable Energy Production"**.

```
#Importing data set - you may copy your code from A3
energy_data3 <- read_excel(
  path = "./Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source.xlsx",
  skip = 12,
  sheet = "Monthly Data",
  col_names=FALSE
```

```r
)
read_col_names<- read_excel(
  path = "./Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source.xlsx", skip = 10, n_ma
)
colnames(energy_data3) <- read_col_names
data3 <- energy_data3[,c(1,5)]
nobs <- nrow(data3)
tsdata3 <- ts(data3[,2], frequency=12,start=c(1973,1))
```

Stochastic Trend and Stationarity Tests

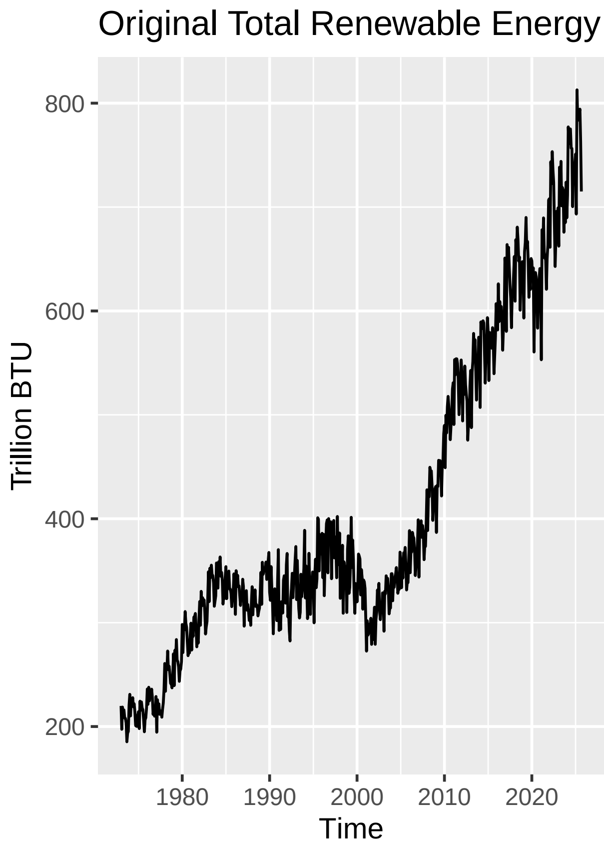
For this part you will work only with the column Total Renewable Energy Production.

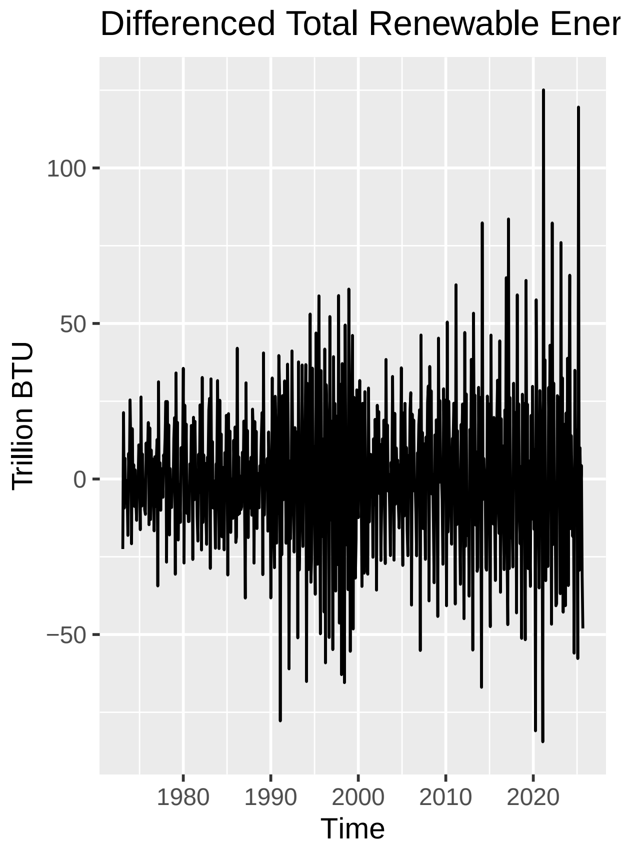
Q1

Difference the "Total Renewable Energy Production" series using function diff(). Function diff() is from package base and take three main arguments: * *x* vector containing values to be differenced; * *lag* integer indicating with lag to use; * *differences* integer indicating how many times series should be differenced.

Try differencing at lag 1 only once, i.e., make `lag=1` and `differences=1`. Plot the differenced series. Do the series still seem to have trend?

```r
diff_renewables <- diff(tsdata3, lag = 1, differences = 1)
diff_renewables_plot <- autoplot(diff_renewables) +
  labs(title = "Differenced Total Renewable Energy Production",
    x = "Time",
    y = "Trillion BTU"
  )
original_renewables_plot <- autoplot(tsdata3) +
  labs(title = "Original Total Renewable Energy Production",
    x = "Time",
    y = "Trillion BTU"
  )
plot_grid(original_renewables_plot, diff_renewables_plot, nrow = 1, align = "h")
```





> Answer: No, there is no clear trend from the plot as the values fluctuates around 0, showing more of a regular variation rather than a general upward trend like the original series.

Q2

Copy and paste part of your code for A3 where you run the regression for Total Renewable Energy Production and subtract that from the original series. This should be the code for Q3 and Q4. make sure you use assign same name for the time series object that you had in A3, otherwise the code will not work.

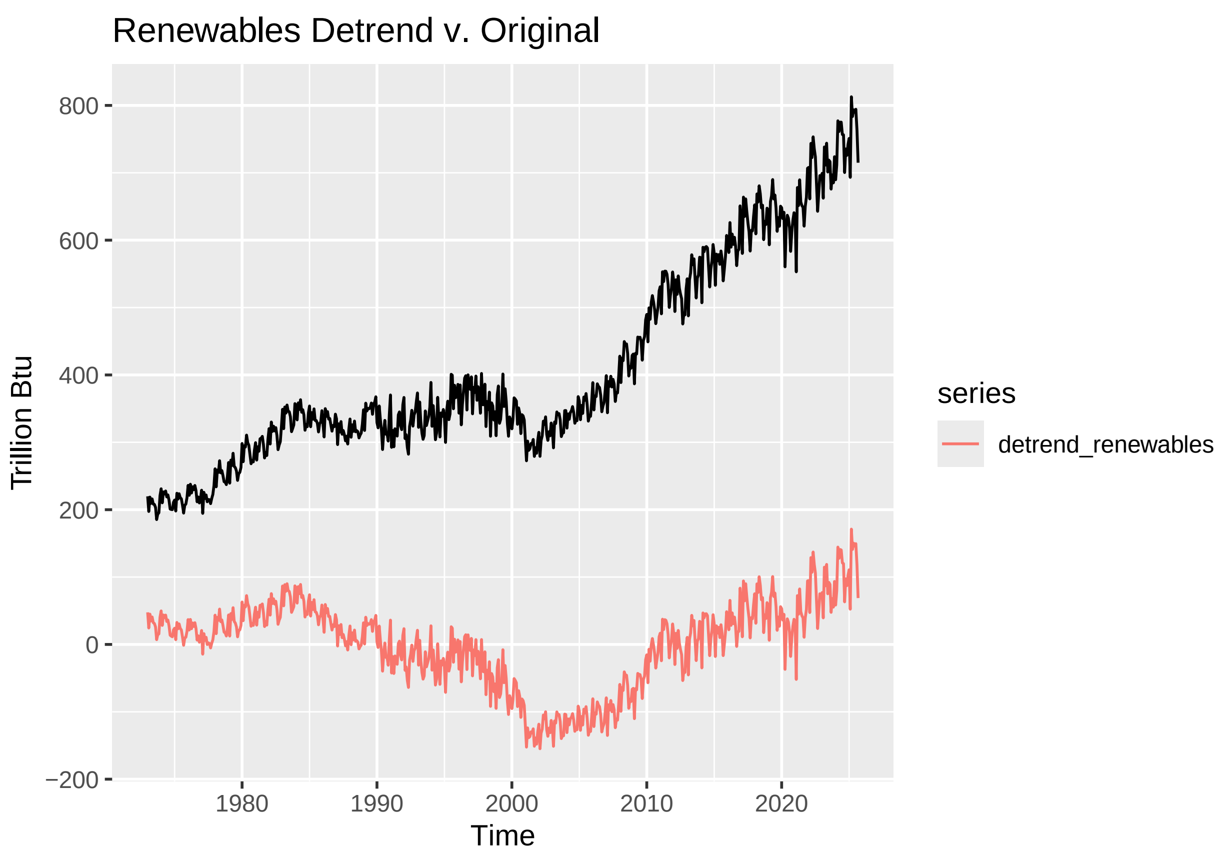
```
t <- c(1:nobs)
renewables_linear_trend_model <- lm(tsdata3 ~ t)
summary(renewables_linear_trend_model)
```

```
##
## Call:
## lm(formula = tsdata3 ~ 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
```

```
renewables_beta0 <- as.numeric(renewables_linear_trend_model$coefficients[1]) #intercept
renewables_beta1 <- as.numeric(renewables_linear_trend_model$coefficients[2]) #slope

detrend_renewables <- tsdata3 - (renewables_beta0 + renewables_beta1*t)
autoplot(tsdata3) +
autolayer(detrend_renewables) +
labs(x = "Time", y = "Trillion Btu", title = "Renewables Detrend v. Original")
```



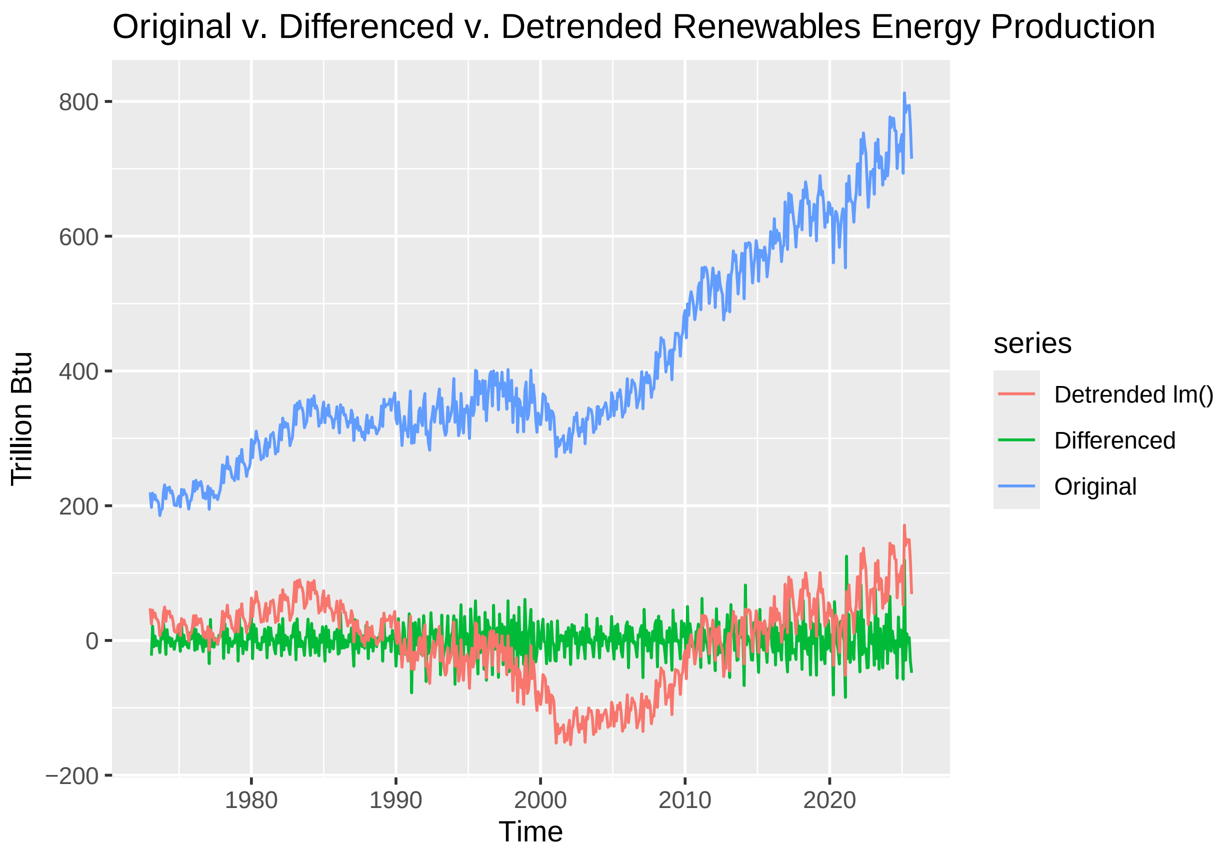
Q3

Now let's compare the differenced series with the detrended series you calculated on A3. In other words, for the "Total Renewable Energy Production" compare the differenced series from Q1 with the series you detrended in Q2 using linear regression.

Using autoplot() + autolayer() create a plot that shows the three series together (i.e. "Original", "Differenced", "Detrended lm()"). Make sure your plot has a legend. The easiest way to do it is by adding the `series=` argument to each autoplot and autolayer function. Look at the key for A03 for an example on how to use autoplot() and autolayer().

What can you tell from this plot? Which method seems to have been more efficient in removing the trend?

```
autoplot(tsdata3, series = "Original") +
  autolayer(diff_renewables, series = "Differenced") +
  autolayer(detrend_renewables, series = "Detrended lm()") +
  labs(x = "Time", y = "Trillion Btu", title = "Original v. Differenced v. Detrended Renewables Energy P
```



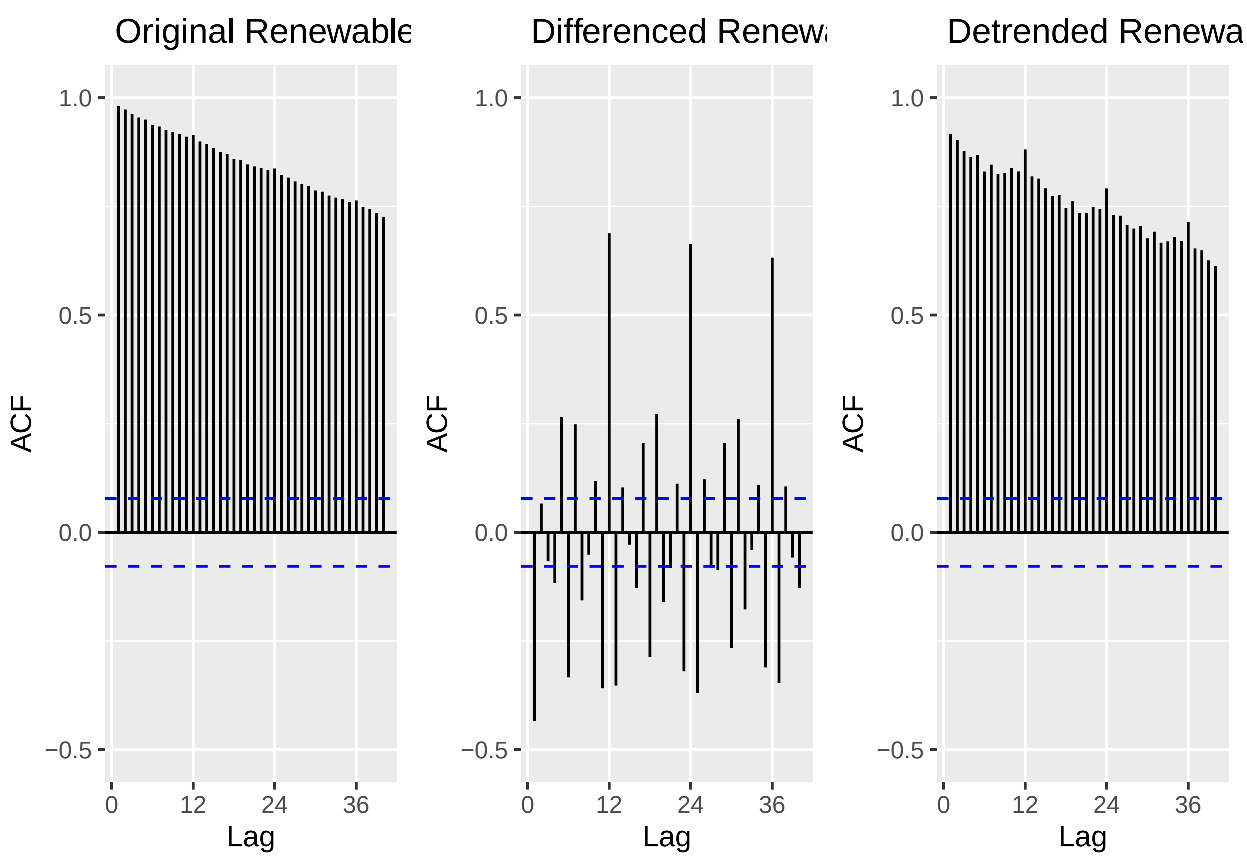
Answer: The original plot shows a clear upward trend. The detrended plot shows a decreasing trend and then increasing trend when linear slope is removed. The differenced plot is relatively flat and shows fluctuates tightly around zero. It doesn't show any increasing or decreasing trends. Differencing seems to be the most efficient method in terms of removing the trend because it focuses more on the local variance, thus being a more effective way to achieve stationarity.

Q4

Plot the ACF for the three series and compare the plots. Add the argument `ylim=c(-0.5,1)` to the autoplot() or Acf() function - whichever you are using to generate the plots - to make sure all three y axis have the same limits. Looking at the ACF which method do you think was more efficient in eliminating the trend? The linear regression or differencing?

```
original_renewables_acf <- Acf(tsdata3, lag.max = 40, plot = FALSE) %>%
  autoplot() +
  labs(title = " Original Renewables ACF", x = "Lag", y = "ACF") +
  ylim(-0.5, 1) +
  theme(
    axis.ticks.x = element_line(),
```

```r
    axis.text.x = element_text()
  )
diff_renewables_acf <- Acf(diff_renewables, lag.max = 40, plot = FALSE) %>%
  autoplot() +
  labs(title = " Differenced Renewables ACF", x = "Lag", y = "ACF") +
  ylim(-0.5, 1) +
  theme(
    axis.ticks.x = element_line(),
    axis.text.x = element_text()
  )
detrended_renewables_acf <- Acf(detrend_renewables, lag.max = 40, plot = FALSE) %>%
  autoplot() +
  labs(title = " Detrended Renewables ACF", x = "Lag", y = "ACF") +
  ylim(-0.5, 1) +
  theme(
    axis.ticks.x = element_line(),
    axis.text.x = element_text()
  )
plot_grid(original_renewables_acf, diff_renewables_acf, detrended_renewables_acf, nrow = 1)
```



> Answer: The ACF plots show that differencing is still the most efficient method in eliminating the trend. The differenced ACF plot show large and regular variations between lags, indicating that most of the long-term dependence caused by the trend has been removed.

Q5

Compute the Seasonal Mann-Kendall and ADF Test for the original "Total Renewable Energy Production" series. Ask R to print the results. Interpret the results for both test. What is the conclusion from the Seasonal Mann Kendall test? What's the conclusion for the ADF test? Do they match what you observed in Q3 plot? Recall that having a unit root means the series has a stochastic trend. And when a series has stochastic trend we need to use differencing to remove the trend.

```
smk_result <- SeasonalMannKendall(tsdata3)
print("Results for Seasonal Mann Kendall")
```

```
## [1] "Results for Seasonal Mann Kendall"
```

```
print(smk_result)
```

```
## tau = 0.799, 2-sided pvalue =< 2.22e-16
```

```
print("Results for ADF test")
```

```
## [1] "Results for ADF test"
```

```
print(adf.test(tsdata3))
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  tsdata3
## Dickey-Fuller = -1.0247, Lag order = 8, p-value = 0.9347
## alternative hypothesis: stationary
```

> Answer: The conclusion from the SMK test shows a very small p-value (<0.05), which indicates that we reject the null hypothesis and that the orginial series indeed has monotonic trend. The conclusion for the ADF test shows a large p-value (>0.05), which means we failed to reject the null hypothesis and that the original series has a unit root (non-stationary). Both conclusions support what I observed in Q3 plot, which is that the original series displays a clear upward trend. It contains a unit root and has a stochastic trend.

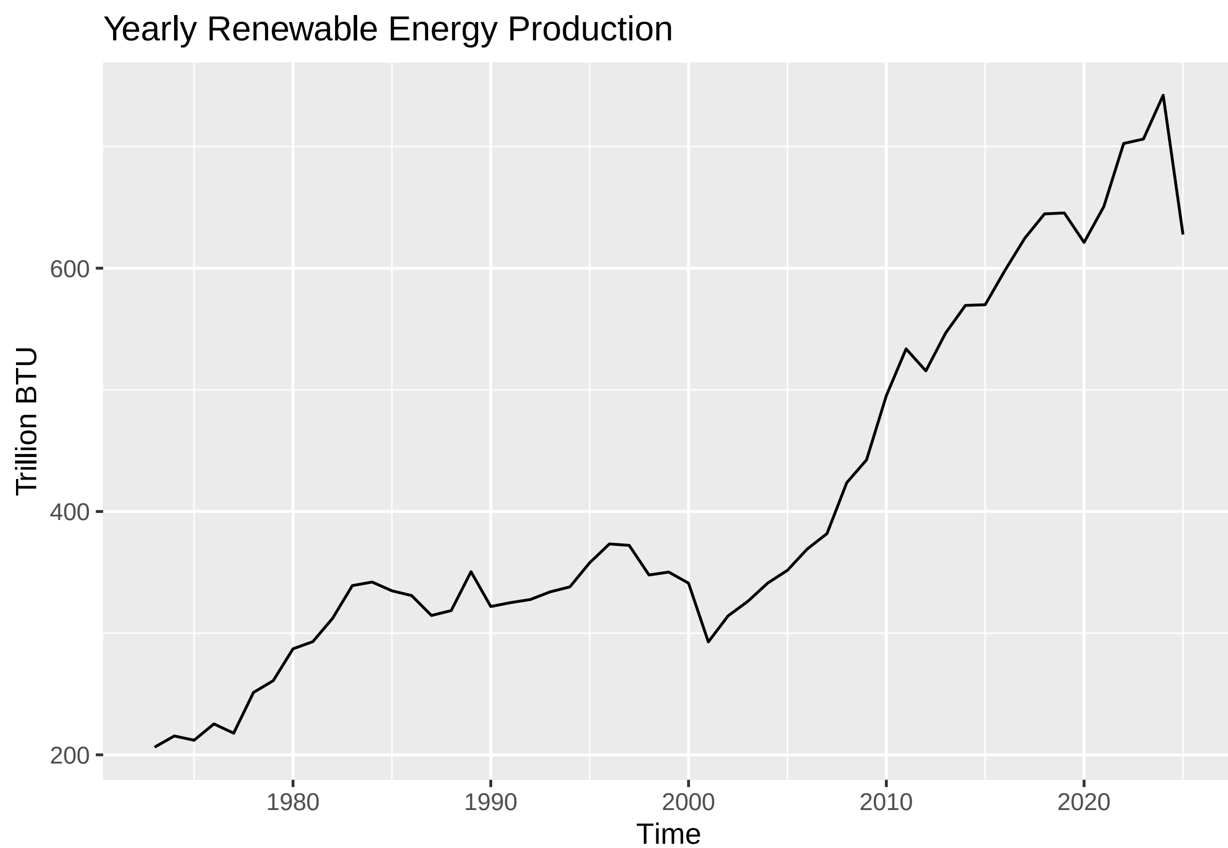
Q6

Aggregate the original "Total Renewable Energy Production" series by year. You can use the same procedure we used in class. Store series in a matrix where rows represent months and columns represent years. And then take the columns mean using function colMeans(). Recall the goal is the remove the seasonal variation from the series to check for trend. Convert the accumulates yearly series into a time series object and plot the series using autoplot().

```
renewables_vector <- as.numeric(tsdata3)
renewables_data_matrix <- matrix(renewables_vector,byrow=FALSE,nrow=12)
```

```
## Warning in matrix(renewables_vector, byrow = FALSE, nrow = 12): data length
## [633] is not a sub-multiple or multiple of the number of rows [12]
```

```r
renewables_data_yearly_means <- colMeans(renewables_data_matrix)
renewables_yearly_ts <- ts(renewables_data_yearly_means, start = 1973, frequency = 1)
autoplot(renewables_yearly_ts) +
  labs(title = "Yearly Renewable Energy Production",
    x = "Time",
    y = "Trillion BTU"
  )
```



Q7

Apply the Mann Kendall, Spearman correlation rank test and ADF. Are the results from the test in agreement with the test results for the monthly series, i.e., results for Q5?

```r
mk_yearly <- MannKendall(renewables_yearly_ts)
print("Results for Yearly Mann Kendall")
```

```
## [1] "Results for Yearly Mann Kendall"
```

```r
print(mk_yearly)
```

```
## tau = 0.816, 2-sided pvalue =< 2.22e-16
```

```r
years <- as.numeric(time(renewables_yearly_ts))
spearman_results <- cor(renewables_yearly_ts,years,method="spearman")
print("Results for Yearly Spearman Test")
```

```
## [1] "Results for Yearly Spearman Test"
```

```r
print(mk_yearly)
```

```
## tau = 0.816, 2-sided pvalue =< 2.22e-16
```

```r
print("Results for ADF test on yearly data")
```

```
## [1] "Results for ADF test on yearly data"
```

```r
print(adf.test(renewables_yearly_ts, alternative = "stationary"))
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  renewables_yearly_ts
## Dickey-Fuller = -1.6789, Lag order = 3, p-value = 0.7037
## alternative hypothesis: stationary
```

> Answer: The Mann Kendall, Spearman test, and ADF all match with the results from the tests results for the monthly series. The Mann Kendall for yearly shows a small p-value, which means we reject the null hypothesis. From the plot we can also see that there is clear upward trend. The Spearman test show the same result. The ADF result shows a large p-value, which means we fail to reject the null hypothesis and that the series has a unit root and is non-stationary.