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

Assignment 4 - Due date 02/10/26

Judy Hua Zhu

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(ggplot2)
library(forecast)
library(readxl)
library(tseries)
library(Kendall)
library(trend)
library(here)
here()
```

```
## [1] "/home/guest/TSA_Sp26"
```

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"**.

```r
#Importing data set - you may copy your code from A3
file_path <- here("Data", "Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source.xlsx")
energy_data1 <- read_excel(file_path,skip = 12, sheet="Monthly Data",col_names=FALSE)
read_col_names <- read_excel(file_path,skip = 10,n_max = 1, sheet="Monthly Data",col_names=FALSE)
#Assign the column names to the data set
colnames(energy_data1) <- read_col_names

renewable <- energy_data1[,5]
renewable_ts <- ts(renewable, start=c(1973,1), frequency=12)
```

Stochastic Trend and Stationarity Tests

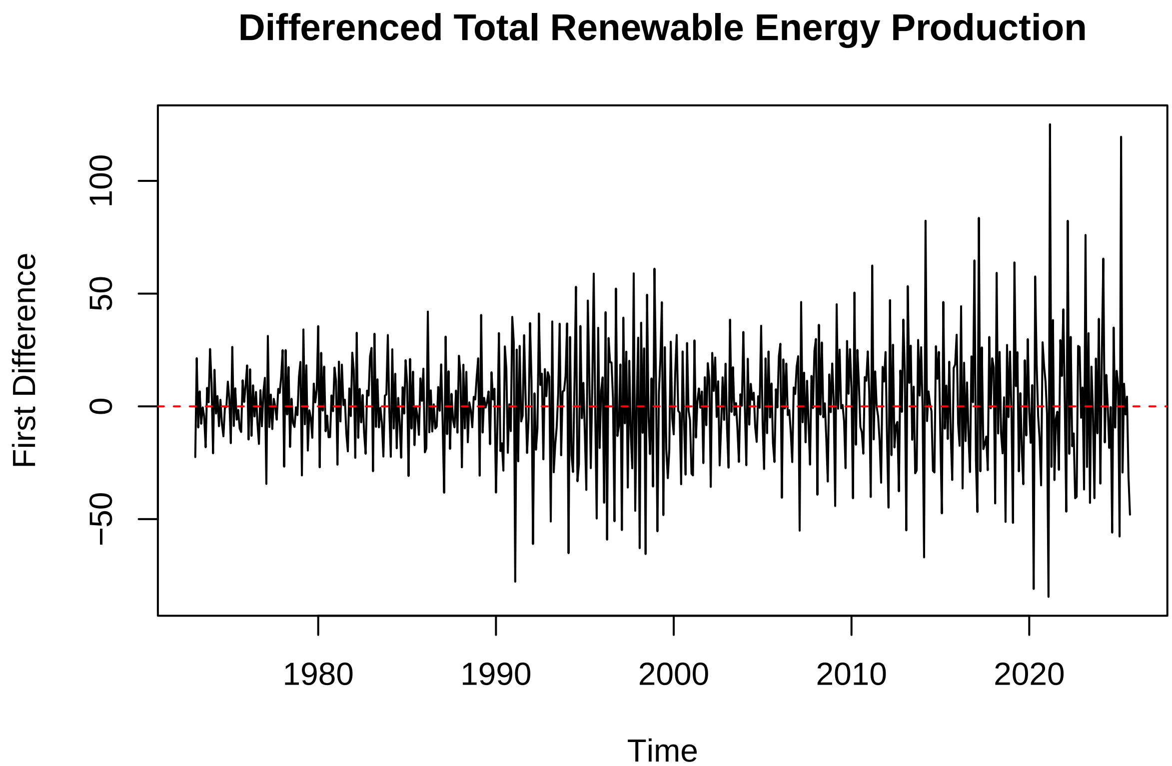
For this part you will work only with the column Total Renewable Energy Production.

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_series <- diff(
  renewable_ts,
  lag = 1,
  differences = 1
)
plot(
  diff_series,
  type = "l",
  main = "Differenced Total Renewable Energy Production",
  ylab = "First Difference",
  xlab = "Time"
)
abline(h = 0, col = "red", lty = 2)
```



No, the series does not show a clear trend.

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.

```r
#Create vector t
row <- nrow(renewable)
t <- c(1:row)
renewables_lm <- lm(renewable_ts~t)
summary(renewables_lm)
```

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

```r
#Store reg coefficient
re_0 <- as.numeric(renewables_lm$coefficients[1]) #intercept
re_1 <- as.numeric(renewables_lm$coefficients[2]) #slope
detrend_renewable <- renewable_ts - (re_0 + re_1*t)
```

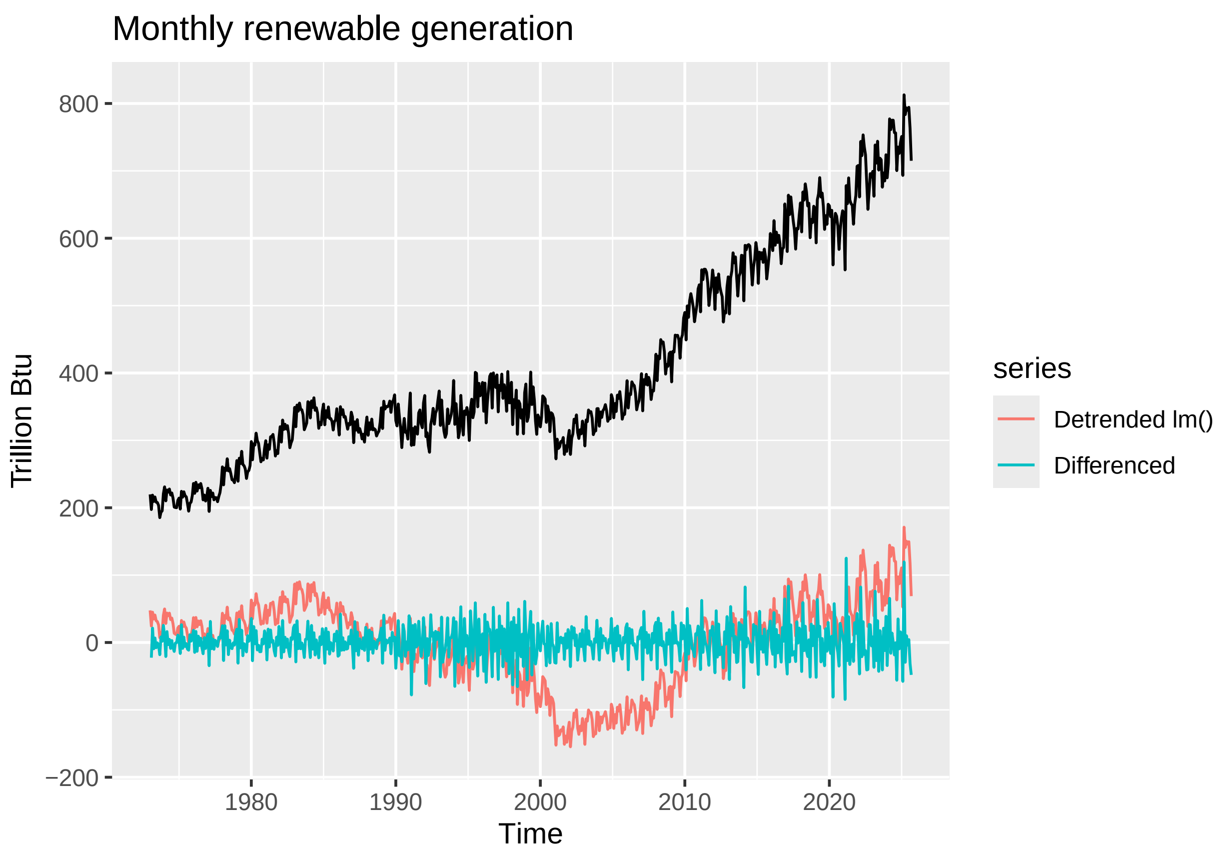
Q3

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?

```r
autoplot(renewable_ts)+
  autolayer(detrend_renewable, series = "Detrended lm()")+
  autolayer(diff_series, series = "Differenced")+
  labs(title = "Monthly renewable generation", x = "Time", y = "Trillion Btu")
```



Answer: Differencing removes the trend in the series by converting the data from levels to period-to-period changes, resulting in a series that fluctuates around a constant mean (0). It seems to be more efficient in removing the trend.

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_ACF <- autoplot(
  Acf(renewable_ts, plot = FALSE),
  ylim = c(-0.5, 1)
) + ggtitle("ACF: Original Series")
```

```
## Warning in ggplot2::geom_segment(lineend = "butt", ...): Ignoring unknown
## parameters: 'ylim'
```

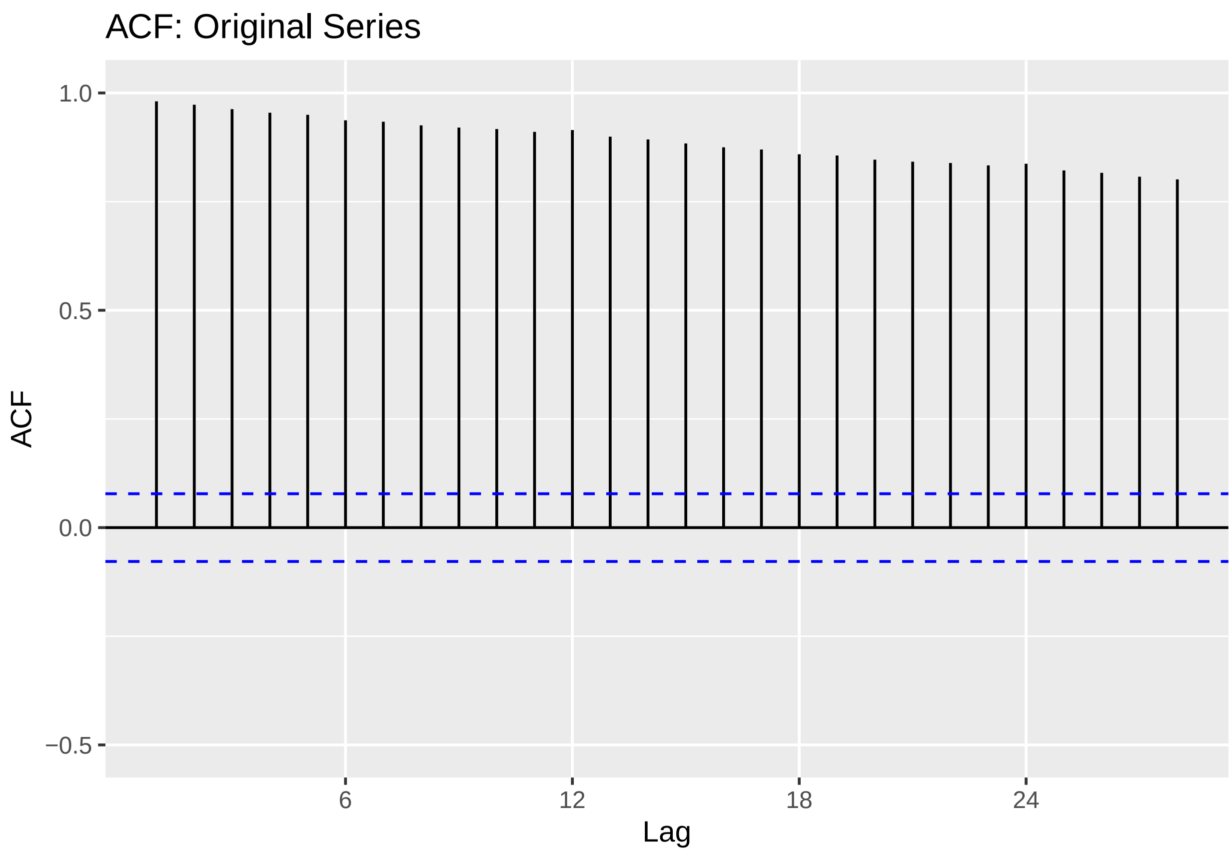
```
detrend_ACF <- autoplot(
  Acf(detrend_renewable, plot = FALSE),
  ylim = c(-0.5, 1)
) + ggtitle("ACF: Detrended (lm)")
```

```
## Warning in ggplot2::geom_segment(lineend = "butt", ...): Ignoring unknown
## parameters: `ylim`
```

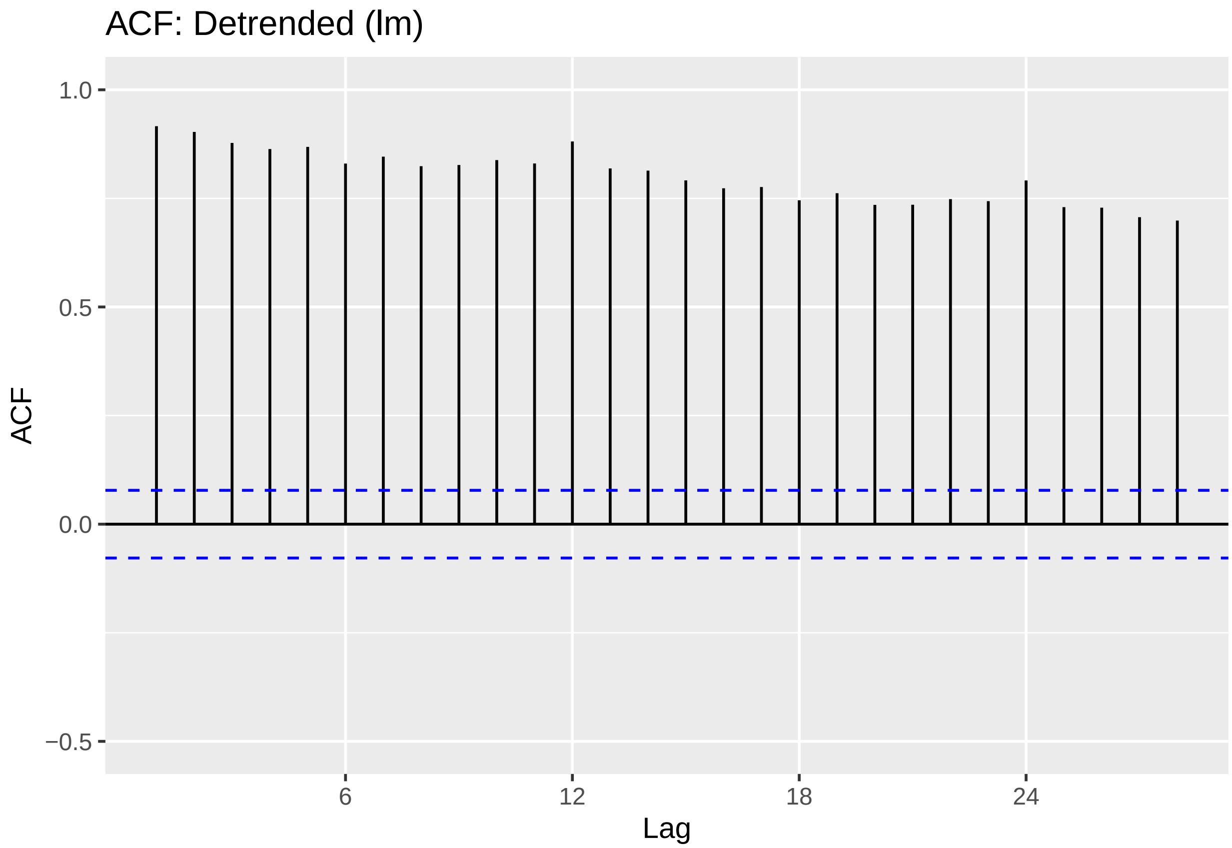
```
differenced_ACF <- autoplot(
  Acf(diff_series, plot = FALSE),
  ylim = c(-0.5, 1)
) + ggtitle("ACF: Differenced")
```

```
## Warning in ggplot2::geom_segment(lineend = "butt", ...): Ignoring unknown
## parameters: `ylim`
```

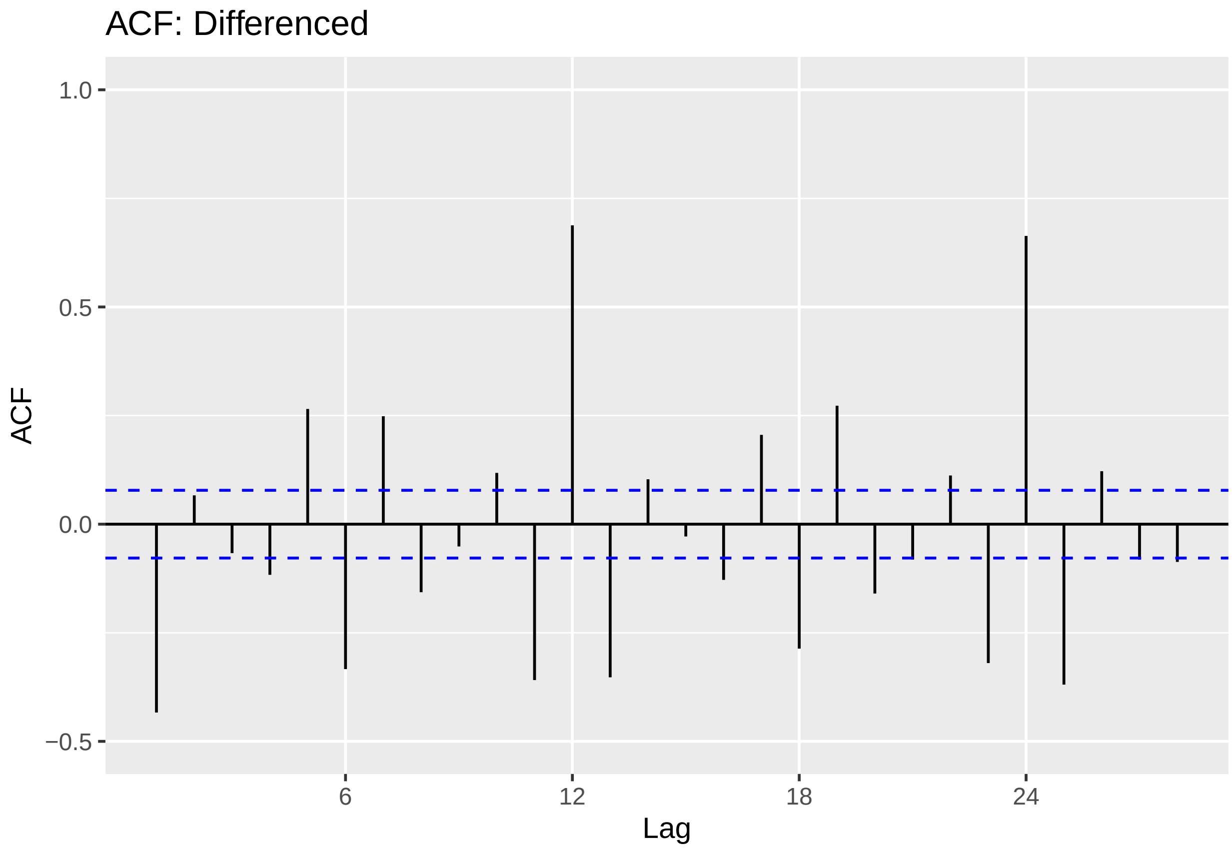
```
original_ACF
```




```
detrend_ACF
```




```
differenced_ACF
```



Answer: Differencing is more efficient in removing as its ACF values fluctuates up and down around 0 with specific significant values. The detrended ACF still shows a positive trend since all values are in positive side even not as constant decreasing as the original series.

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 <- smk.test(renewable_ts)
print(smk_result)
```

```
##
##  Seasonal Mann-Kendall trend test (Hirsch-Slack test)
##
## data:  renewable_ts
## z = 29.17, p-value < 2.2e-16
## alternative hypothesis: true S is not equal to 0
## sample estimates:
##      S   varS
##  13083 201135
```

```r
adf_result <- adf.test(renewable_ts)
print(adf_result)
```

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

> Answer: The p-value for seasonal Mann-Kendall test is lower than 0.05, we reject the null hypothesis, accept the alternative hypothesis, indicating there is a statistically significant increasing trend in Total Renewable Energy Production over time. The p-value for ADF test is greater than 0.05, we fail to reject the null hypothesis of a unit root, indicating that the series is non-stationary and exhibits a stochastic trend. The results of both the Seasonal Mann–Kendall test and the ADF test are consistent with the graphical analysis in Q3. The plot shows a strong upward trend, which is confirmed by the Seasonal Mann–Kendall test, while the ADF test indicates the presence of a unit root and a stochastic trend. This aligns with the observed need to difference the series to achieve stationarity.

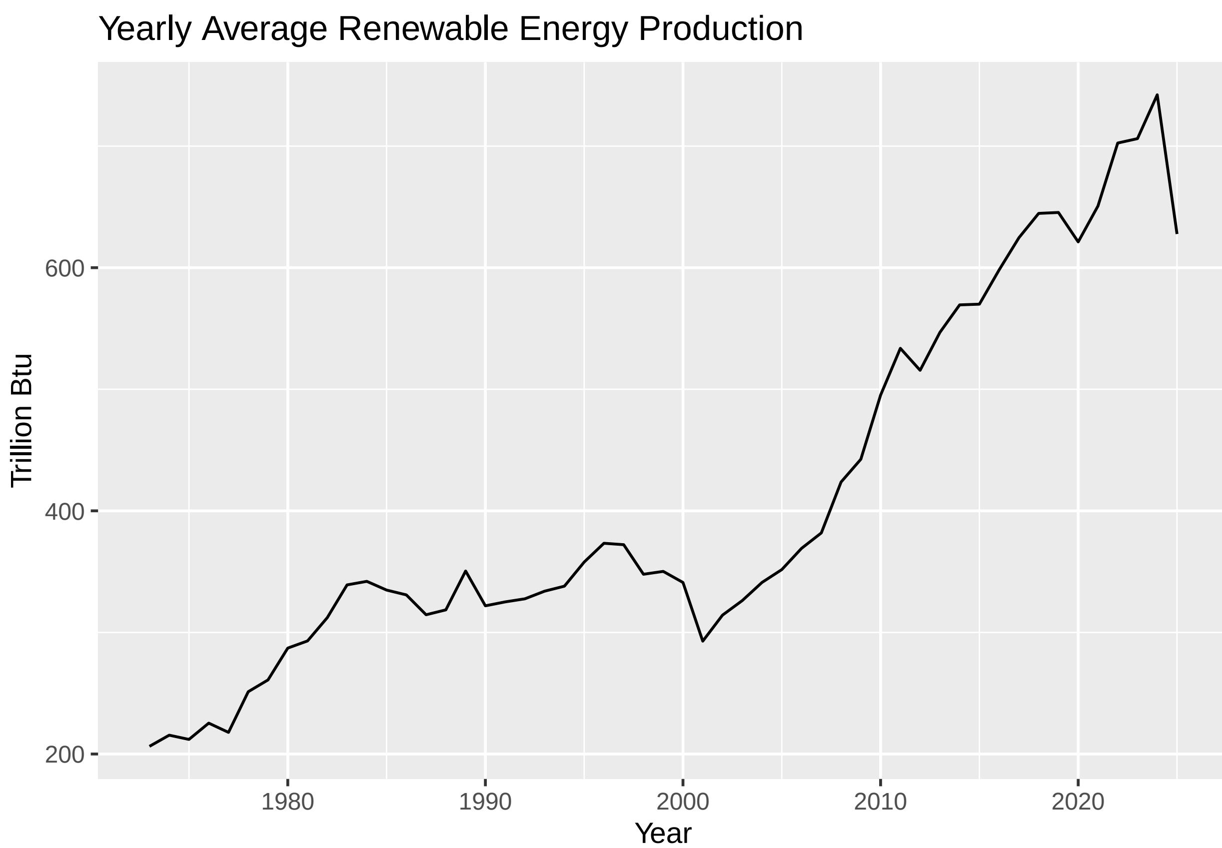
Q6

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().

```r
renewable_matrix <- matrix(
  renewable_ts,
  nrow = 12,
  byrow = FALSE
)
```

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

```r
renewable_yearly <- colMeans(renewable_matrix, na.rm = TRUE)
renewable_yearly_ts <- ts(
  renewable_yearly,
  start = c(1973),
  frequency = 1
)
autoplot(renewable_yearly_ts) +
  labs(
    title = "Yearly Average Renewable Energy Production",
    x = "Year",
    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?

```
mk_yearly <- mk.test(renewable_yearly_ts)
print(mk_yearly)
```

```
##
##  Mann-Kendall trend test
##
## data:  renewable_yearly_ts
## z = 8.6142, n = 53, p-value < 2.2e-16
## alternative hypothesis: true S is not equal to 0
## sample estimates:
##            S         varS          tau
## 1.124000e+03 1.699533e+04 8.156749e-01
```

```
time_index <- time(renewable_yearly_ts)

spearman_yearly <- cor.test(
  renewable_yearly_ts,
  time_index,
  method = "spearman"
```

```
)
print(spearman_yearly)
```

```
##
##  Spearman's rank correlation rho
##
## data:  renewable_yearly_ts and time_index
## S = 1898, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##       rho
## 0.9234801
```

```
adf_yearly <- adf.test(renewable_yearly_ts)
print(adf_yearly)
```

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

> Answer: Mann Kendall test shows a results with p-value<0.05 and positive test statistic. We reject the null hypothesis, and conclude there is a statistically significant increasing trend in the yearly renewable energy production series. Spearman's correlation rank test results show a rho of 0.923 and p-value<0.05. We also reject the null hypthesis and conclude there is a strong positive monotonic relationship with time, confirming an increasing trend. The p-value of the ADF test is 0.7>0.05, we fail to reject the null hypothesis of a unit root, indicating that the yearly series is non-stationary and exhibits a stochastic trend. All test results of yearly data are in agreement with the monthly series.