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

Assignment 4 - Due date 02/10/26

Trudy

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(dplyr)
library(forecast)
library(cowplot)
library(ggplot2)
library(forecast)
library(tseries)
library(Kendall)
```

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
file_path <- "./Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source.xlsx"

# Import raw energy data
```

```r
raw_energy_ts <- read_excel(
  path = file_path,
  skip = 12,
  sheet = "Monthly Data",
  col_names = FALSE
)

# Extract header row
header_row <- read_excel(
  path = file_path,
  skip = 10,
  n_max = 1,
  sheet = "Monthly Data",
  col_names = FALSE
)

# Assign column names
colnames(raw_energy_ts) <- as.character(header_row[1, ])

# Keep only required column
renewable_df <- raw_energy_ts %>%
  select(
    Date = 1,
    `Total Renewable Energy Production`
  )

# Preview
head(renewable_df)
```

```
## # A tibble: 6 x 2
##   Date                `Total Renewable Energy Production`
##   <dttm>                                             <dbl>
## 1 1973-01-01 00:00:00                                 220.
## 2 1973-02-01 00:00:00                                 197.
## 3 1973-03-01 00:00:00                                 219.
## 4 1973-04-01 00:00:00                                 209.
## 5 1973-05-01 00:00:00                                 216.
## 6 1973-06-01 00:00:00                                 208.
```

Stochastic Trend and Stationarity Tests

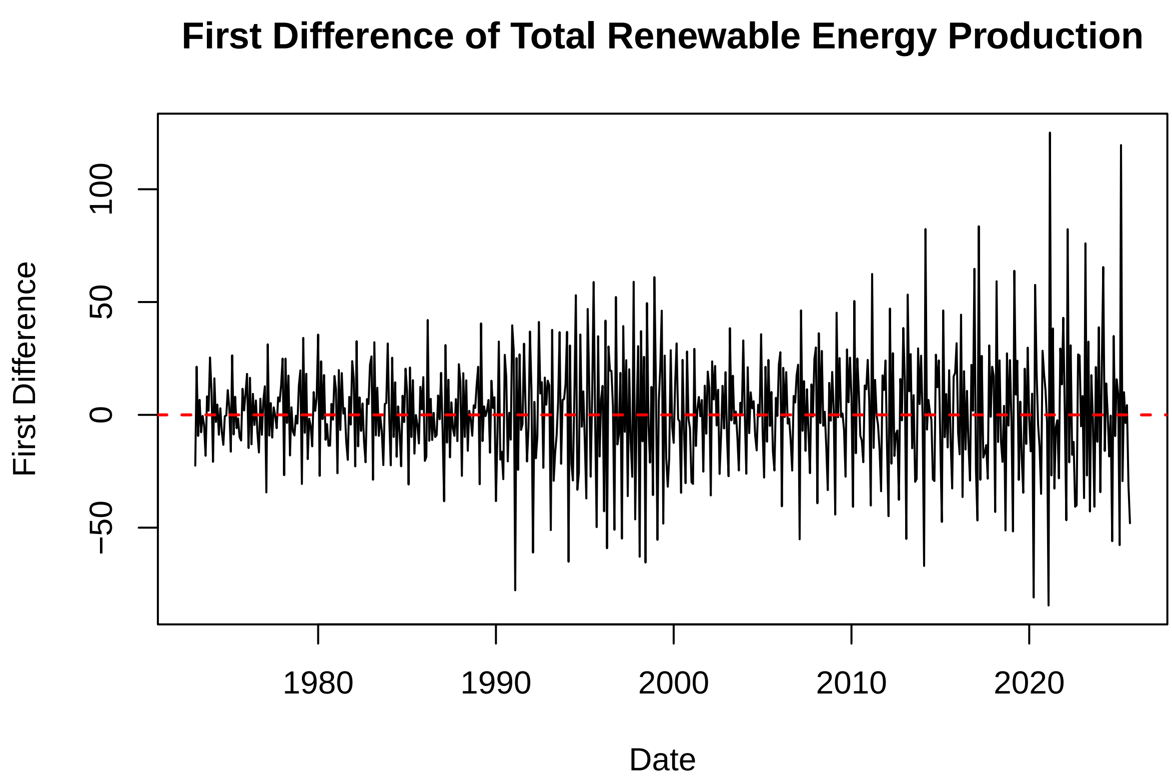
For this part you will work only with the column Total Renewable Energy Production.

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_df$`Total Renewable Energy Production`,
  lag = 1,
  differences = 1
)

diff_df <- data.frame(
  Date = renewable_df$Date[-1],
  Diff_Value = diff_series
)

plot(
  diff_df$Date,
  diff_df$Diff_Value,
  type = "l",
  main = "First Difference of Total Renewable Energy Production",
  xlab = "Date",
  ylab = "First Difference"
)

abline(h = 0, col = "red", lty = 2, lwd = 1.5)
```



Answer: The first differenced series no longer shows a clear trend and fluctuates around a constant mean near zero, indicating that differencing successfully removed the stochastic trend. However, the variability appears somewhat larger in later years, indicating that variance may not be perfectly constant.

Q2

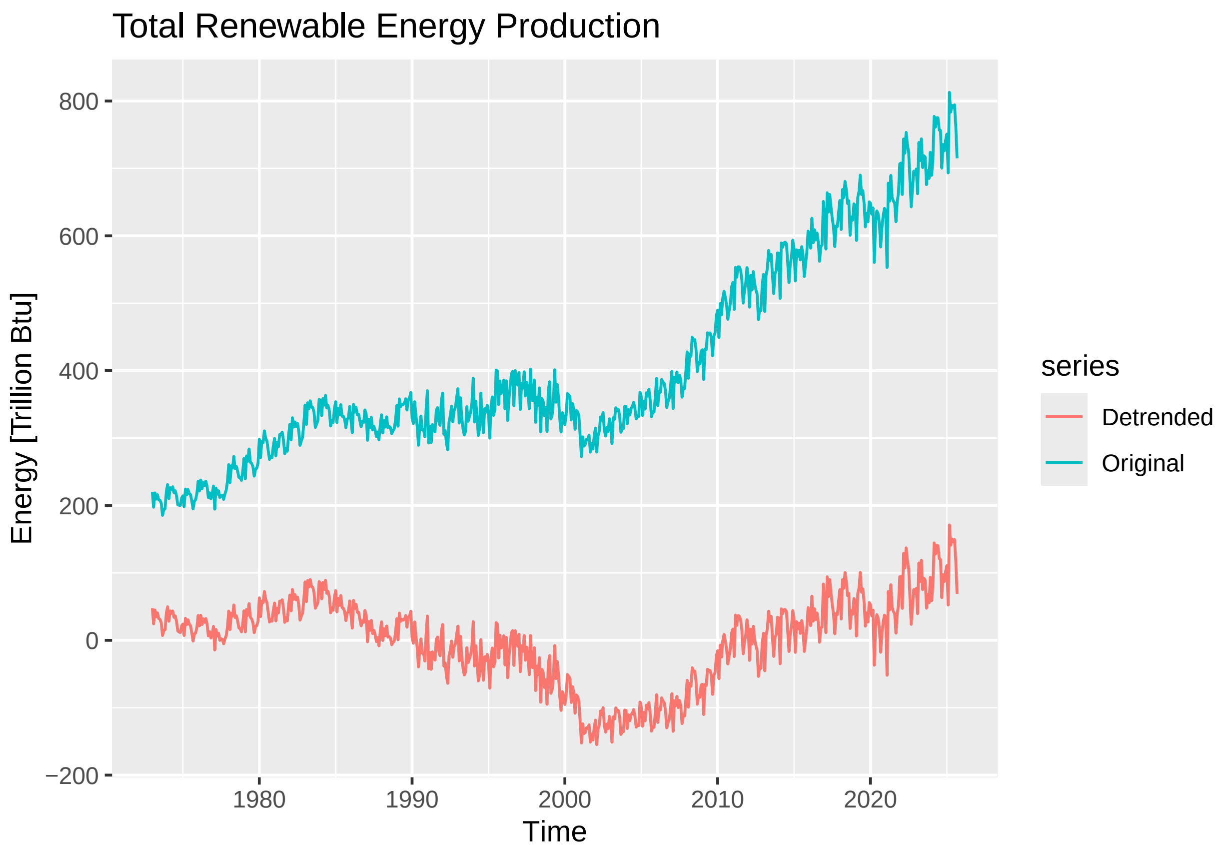
Copy and paste part of your code for A3 where you run the regression for Total Renewable Energy Production and subtract that from the original series. This should be the code for Q3 and Q4. make sure you use assign same name for the time series object that you had in A3, otherwise the code will not work.

```
#create ts object (revise based on A3 answer key)
nobs <- nrow(renewable_df)
t <- 1:nobs

ts_total_renew <- ts(
  renewable_df$`Total Renewable Energy Production`,
  frequency = 12,
  start = c(1973, 1)
)

#Q3_Linear trend regression
regmodel_renewable <- lm(ts_total_renew ~ t)
print(summary(regmodel_renewable))
```

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

```
# Save coefficients
beta0_renewable <- as.numeric(regmodel_renewable$coefficients[1])
beta1_renewable <- as.numeric(regmodel_renewable$coefficients[2])


#Q4_Detrend
renewable_detrend <- ts_total_renew - (beta0_renewable + beta1_renewable * t)

renewable_detrend <- ts(
  renewable_detrend,
  frequency = frequency(ts_total_renew),
```

```r
  start = start(ts_total_renew)
)

# Plot original vs detrended
autoplot(ts_total_renew, series = "Original") +
  autolayer(renewable_detrend, series = "Detrended") +
  ylab("Energy [Trillion Btu]") +
  ggtitle("Total Renewable Energy Production")
```



Q3

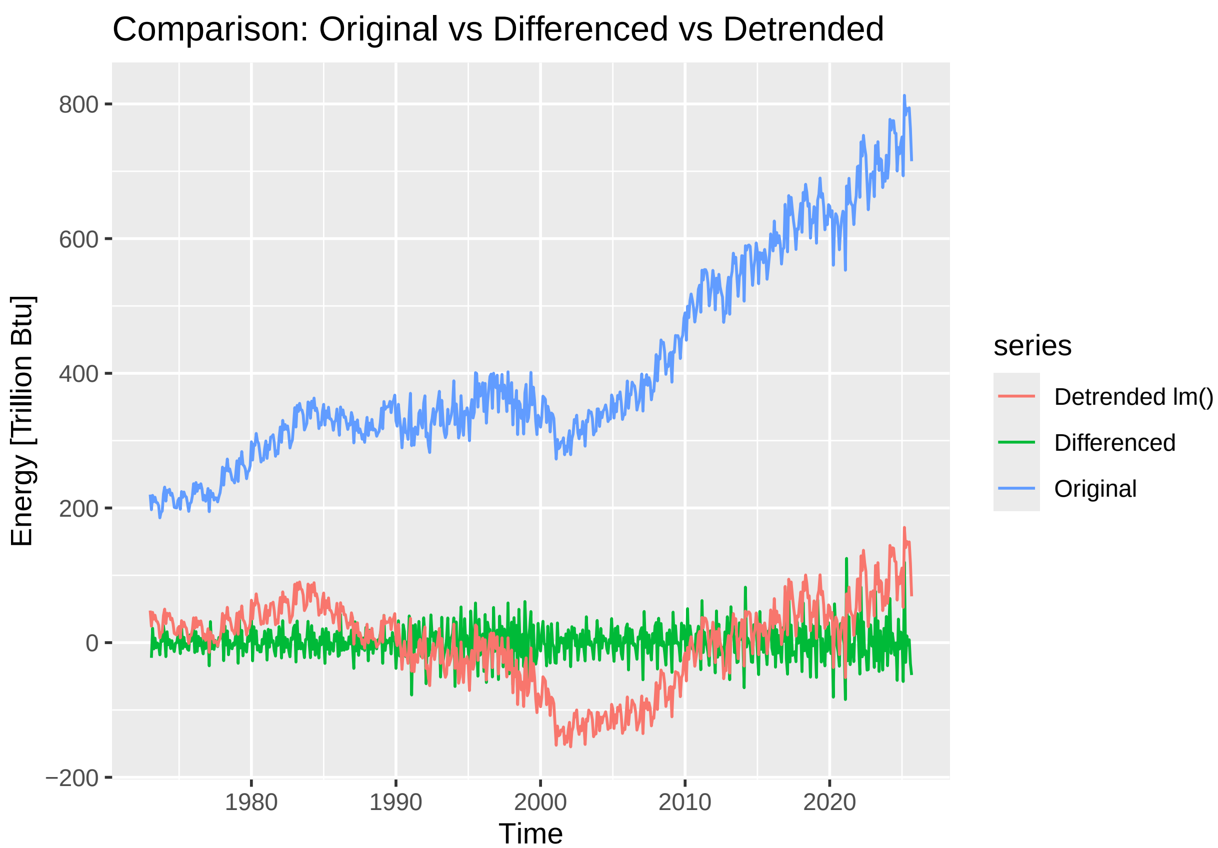
Now let's compare the differenced series with the detrended series you calculated on A3. In other words, for the "Total Renewable Energy Production" compare the differenced series from Q1 with the series you detrended in Q2 using linear regression.

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
ts_diff_renew <- ts(
  diff_series,
  frequency = frequency(ts_total_renew),
```

```r
  start = c(1973, 2)
)

autoplot(ts_total_renew, series = "Original") +
  autolayer(ts_diff_renew, series = "Differenced") +
  autolayer(renewable_detrend, series = "Detrended lm()") +
  ylab("Energy [Trillion Btu]") +
  ggtitle("Comparison: Original vs Differenced vs Detrended")
```



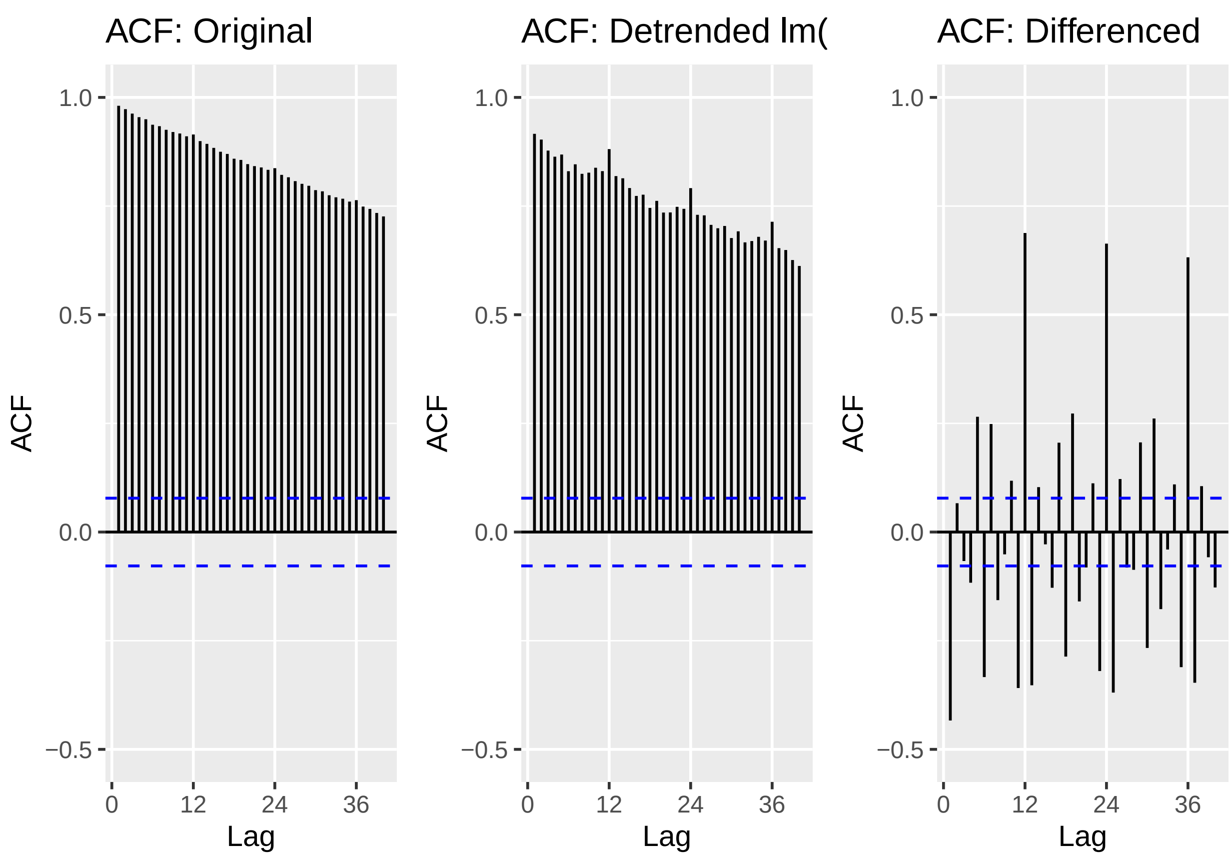
Answer: From the plot we can see that the original series shows a strong upward trend. The detrended series removes part of the linear trend, but some long-term variation still remains. The differenced series fluctuates around zero and does not show any visible trend. Therefore, differencing appears to be more effective than linear detrending in removing the trend and making the series closer to stationary.

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?

```r
acf_original <- autoplot(Acf(ts_total_renew, lag.max = 40, plot = FALSE)) +
  ggtitle("ACF: Original") +
```

```r
  ylim(-0.5, 1)

acf_detrend <- autoplot(Acf(renewable_detrend, lag.max = 40, plot = FALSE)) +
  ggtitle("ACF: Detrended lm()") +
  ylim(-0.5, 1)

acf_diff <- autoplot(Acf(ts_diff_renew, lag.max = 40, plot = FALSE)) +
  ggtitle("ACF: Differenced") +
  ylim(-0.5, 1)

plot_grid(acf_original, acf_detrend, acf_diff, nrow = 1)
```



Answer: Looking at the ACF plots, the original series shows a slow decay in autocorrelation, indicating a strong trend. The detrended series reduces the autocorrelation slightly, but many lags remain significant. The differenced series shows a much faster decay in autocorrelation, with most values close to zero. Therefore, differencing appears to be more efficient than linear regression in eliminating the trend.

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.

```r
#Seasonal Mann Kendall test
smk_test <- SeasonalMannKendall(ts_total_renew)
print(summary(smk_test))
```

```
## Score =  13083 , Var(Score) = 201135
## denominator =  16379.5
## tau = 0.799, 2-sided pvalue =< 2.22e-16
## NULL
```

```r
#ADF test
adf_test <- adf.test(ts_total_renew)
print(adf_test)
```

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

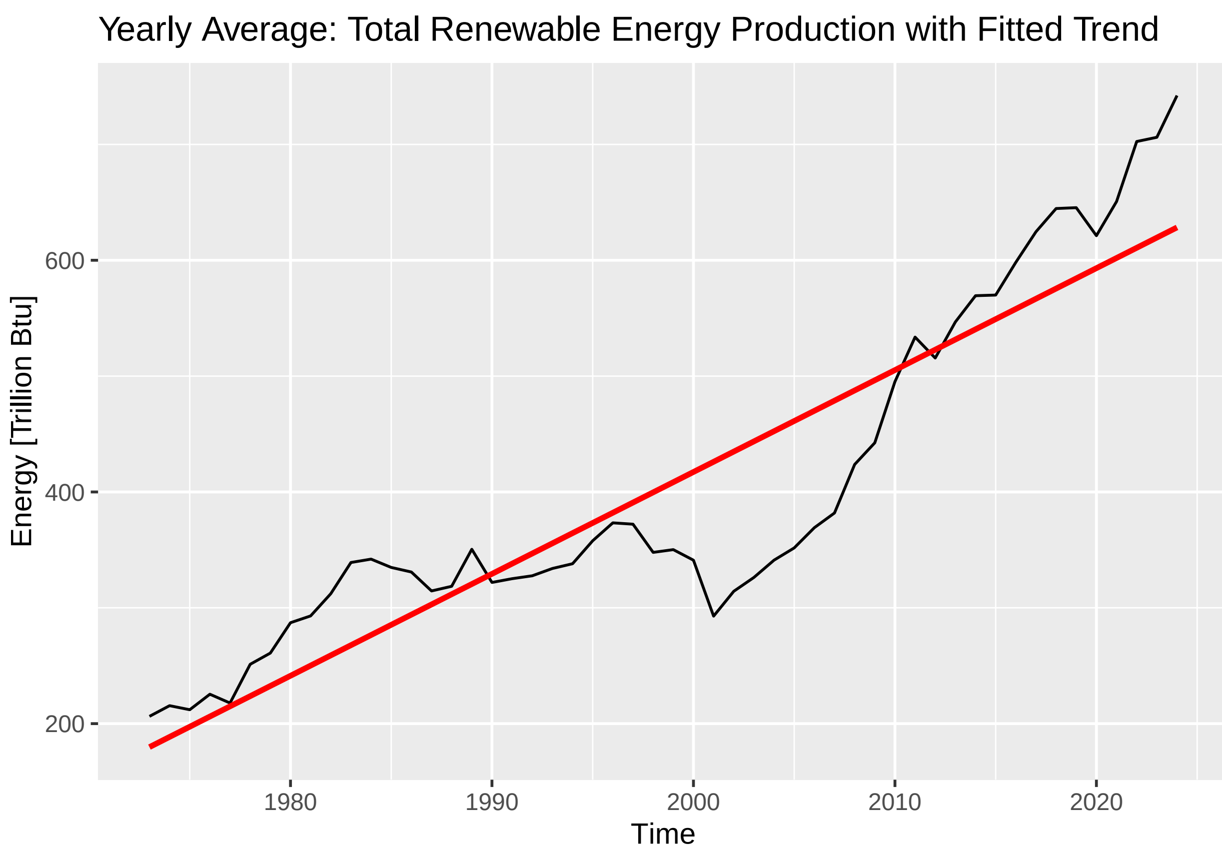
> Answer: The Seasonal Mann-Kendall test indicates a strong and statistically significant upward trend in the original Total Renewable Energy Production series, where tau is 0.799 and p-value is less than 0.05. The ADF test fails to reject the null hypothesis of a unit root, since p-value is 0.9347, indicating that the series is non-stationary and contains a stochastic trend. These results are consistent with the observations from Q3, where the original series showed a clear upward trend and strong persistence in the ACF plot. Since the series has a stochastic trend, differencing is an appropriate method to remove the 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().

```r
# Keep only complete years (multiple of 12 months)
n <- length(ts_total_renew)
n_complete <- floor(n / 12) * 12
ts_trim <- ts_total_renew[1:n_complete]

# Store in matrix
mat_year <- matrix(ts_trim, nrow = 12, byrow = FALSE)

# Yearly averages (removes within-year seasonality)
yearly_mean <- colMeans(mat_year, na.rm = TRUE)

# Convert to yearly ts object and plot
start_year <- start(ts_total_renew)[1]  # e.g., 1973
ts_yearly <- ts(yearly_mean, start = start_year, frequency = 1)
```

```r
# Fit linear trend to yearly series
t_year <- 1:length(ts_yearly)
lm_year <- lm(ts_yearly ~ t_year)

# Fitted trend values
trend_year <- ts(fitted(lm_year),
                 start = start(ts_yearly),
                 frequency = frequency(ts_yearly))

autoplot(ts_yearly, series = "Yearly Average", color = "black") +
  autolayer(trend_year,
            series = "Fitted Trend",
            color = "red",
            linewidth = 1) +
  ylab("Energy [Trillion Btu]") +
  ggtitle("Yearly Average: Total Renewable Energy Production with Fitted Trend")
```



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
#Mann Kendall test
mk_year <- MannKendall(ts_yearly)
print(summary(mk_year))
```

```
## Score =  1084 , Var(Score) = 16059.33
## denominator =  1326
## tau = 0.817, 2-sided pvalue =< 2.22e-16
## NULL
```

```r
#Spearman rank test
spearman_test <- cor.test(
  1:length(ts_yearly),
  ts_yearly,
  method = "spearman"
)
print(spearman_test)
```

```
##
##  Spearman's rank correlation rho
##
## data:  1:length(ts_yearly) and ts_yearly
## S = 1852, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##       rho
## 0.9209425
```

```r
#ADF test
adf_year <- adf.test(ts_yearly)
print(adf_year)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  ts_yearly
## Dickey-Fuller = -0.85301, Lag order = 3, p-value = 0.9515
## alternative hypothesis: stationary
```

> Answer: The Mann-Kendall test indicates a strong and statistically significant upward trend in the yearly aggregated Total Renewable Energy Production series, where tau is 0.817, p-value less than 0.05. The Spearman rank correlation test also shows a strong positive and statistically significant association between time and renewable energy production, where rho is 0.921, p-value less than 0.05, confirming the presence of a monotonic increasing trend. The ADF test fails to reject the null hypothesis of a unit root, where p-value is 0.9515, indicating that the yearly series is non-stationary and contains a stochastic trend. These results are consistent with the results obtained for the monthly series in Q5. Removing the seasonal variation by aggregating the data yearly does not remove the long-term trend, and the series remains non-stationary.