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

Assignment 4 - Due date 02/10/26

Matthew Fischer

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_Consumption
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_data <- read_excel(path="./Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source

read_col_names <- read_excel(path="./Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_Sou

colnames(energy_data) <- read_col_names

energy_subset <- data.frame(Total_Renewable_Energy_Production = energy_data$"Total Renewable Energy Pro
```

```r
head(energy_subset)
```

```
##   Total_Renewable_Energy_Production
## 1                           219.839
## 2                           197.330
## 3                           218.686
## 4                           209.330
## 5                           215.982
## 6                           208.249
```

```r
energy_ts <- ts(
  energy_subset,
  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.

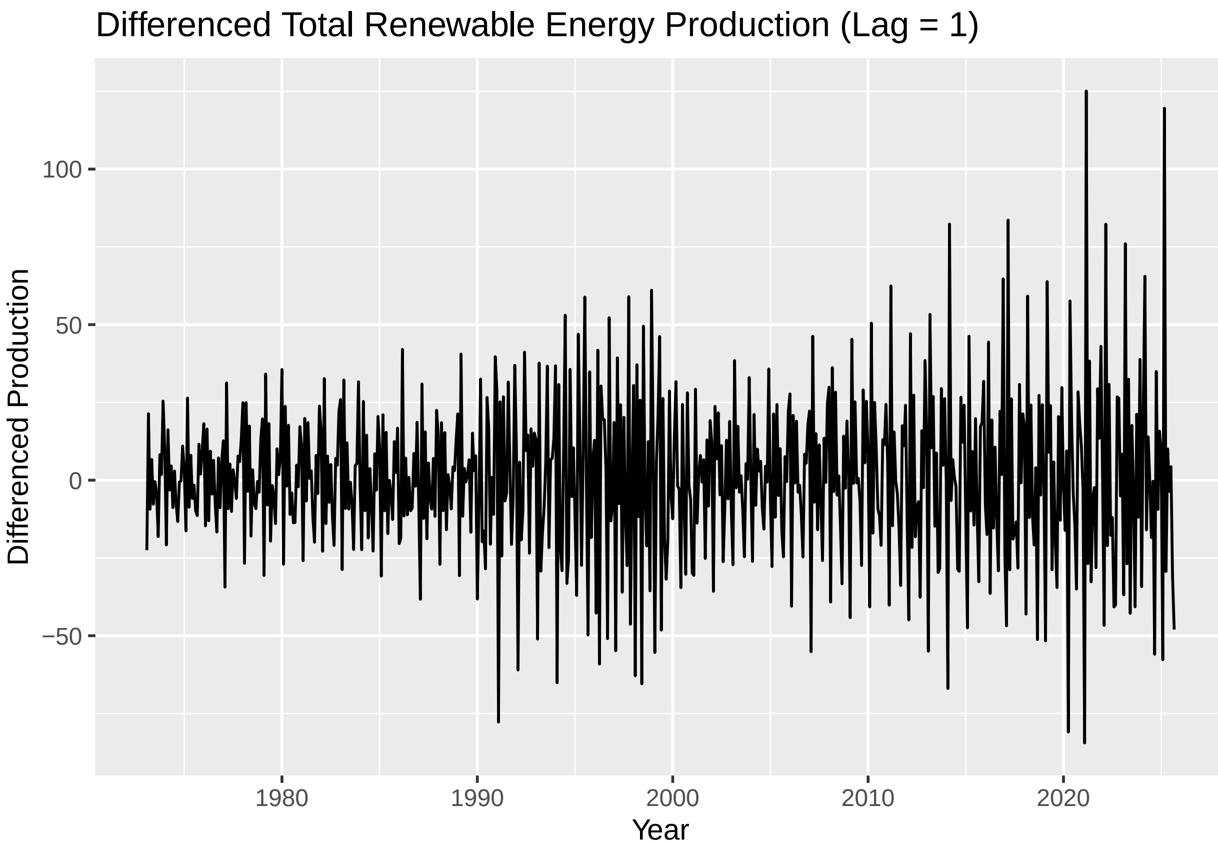
Q1

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
energy_ts_diff <- diff(
  energy_ts,
  lag = 1,
  differences = 1
)

autoplot(energy_ts_diff) +
  ggtitle("Differenced Total Renewable Energy Production (Lag = 1)") +
  xlab("Year") +
  ylab("Differenced Production")
```



Answer: The series no longer seems to have a 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.

```
t <- time(energy_ts)

trp_lm <- lm(energy_ts ~ t)

summary(trp_lm)
```

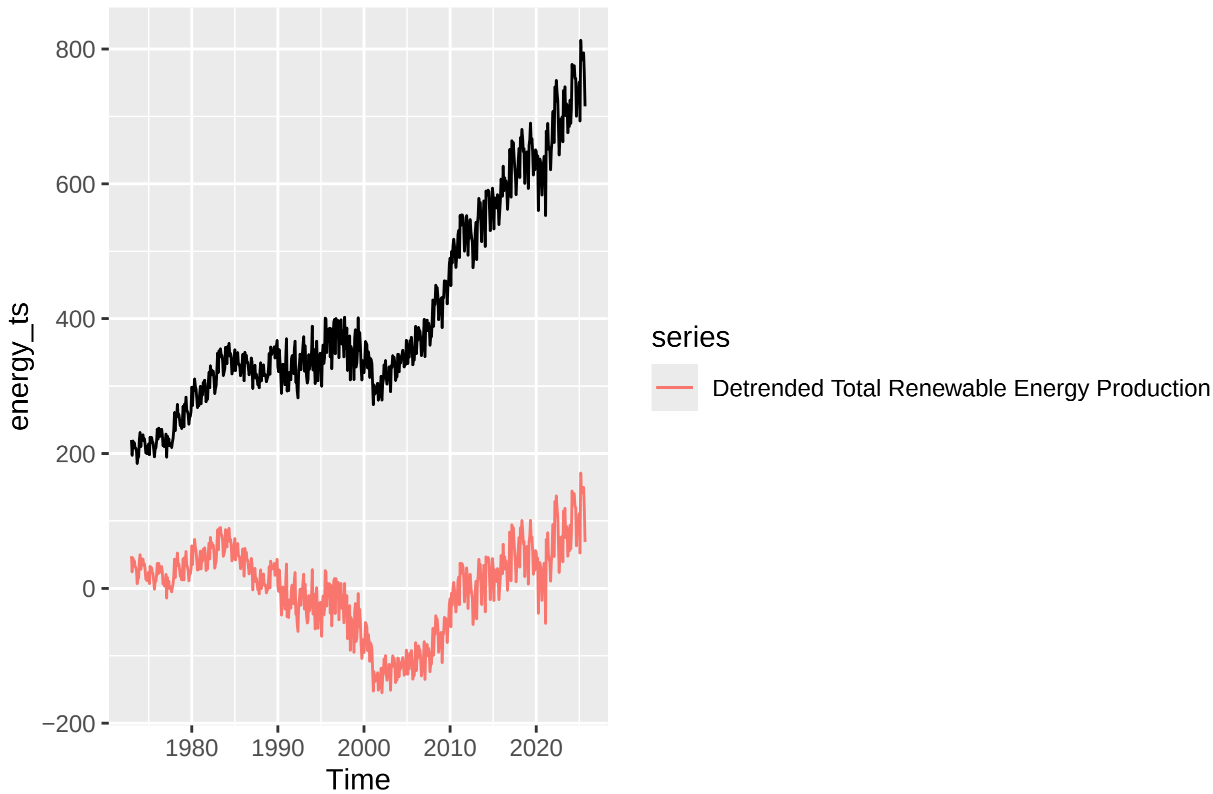
```
##
## Call:
## lm(formula = energy_ts ~ t)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -154.81  -39.55   12.52   41.49  171.15
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.758e+04  3.351e+02  -52.47   <2e-16 ***
## t            9.000e+00  1.676e-01   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
```

```
trp_int <- as.numeric(trp_lm$coefficients[1])
trp_slope <- as.numeric(trp_lm$coefficients[2])

trp_detrended <- energy_ts[, 1] - (trp_int + (trp_slope * t))

autoplot(energy_ts) +
  autolayer(trp_detrended, series = "Detrended Total Renewable Energy Production")
```



Q3

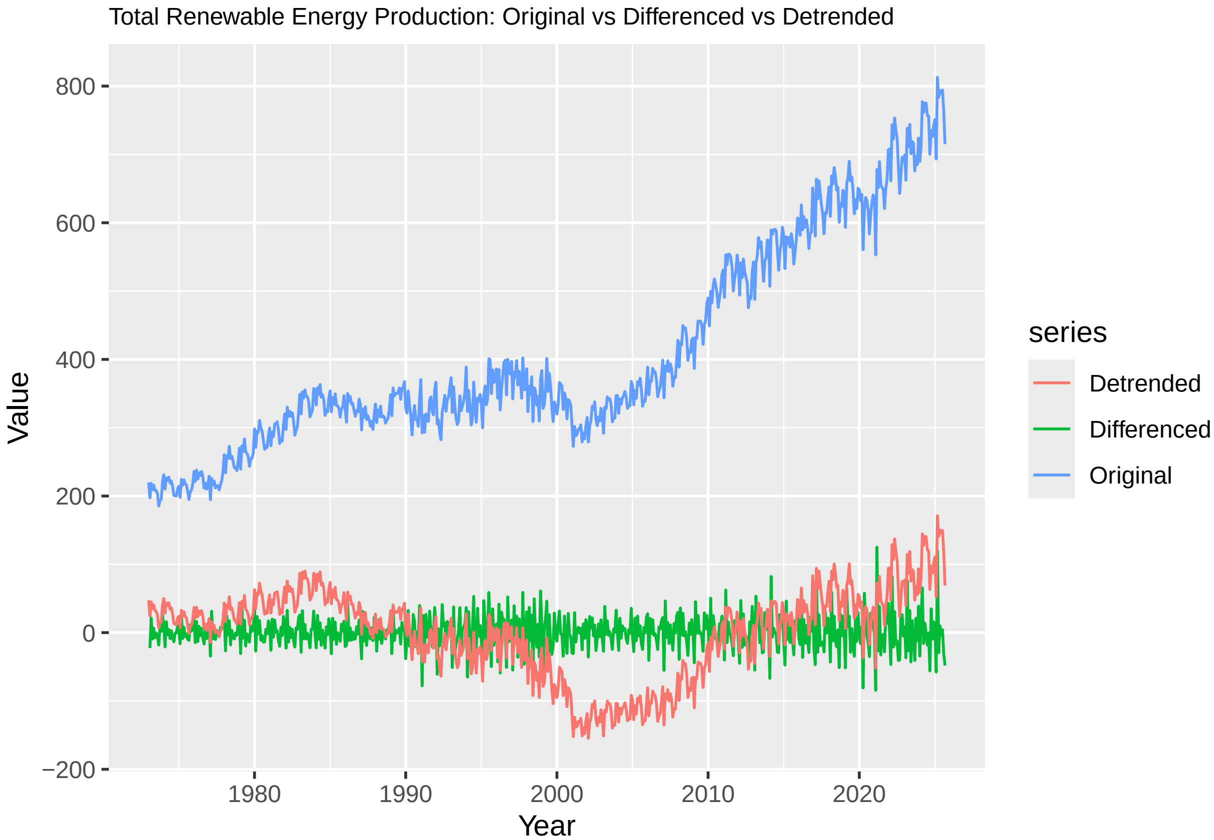
Now let's compare the differenced series with the detrended series you calculated on A3. In other words, for the "Total Renewable Energy Production" compare the differenced series from Q1 with the series you detrended in Q2 using linear regression.

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?

```
autoplot(energy_ts, series = "Original") +
  autolayer(energy_ts_diff, series = "Differenced") +
  autolayer(trp_detrended, series = "Detrended") +
  ggtitle("Total Renewable Energy Production: Original vs Differenced vs Detrended") +
  xlab("Year") +
```

```
  ylab("Value") +
  theme(plot.title = element_text(size = 9))
```



Answer: The detrended and differenced series tend to be at arounf the same height, but the detrended series has larger ovrall increases and decreases while the center of the differenced series seems to always be at 0. Differencing seems to have been the more efficient method 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?

```
trp_acf <- ggAcf(energy_ts, lag.max = 40, ylim = c(-0.5, 1)) +
  ggtitle(paste("Total Renewable Energy Production ACF")) +
  theme_minimal() +
  theme(plot.title = element_text(size = 5))
```

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

```
trp_detrended_acf <- ggAcf(trp_detrended, lag.max = 40, ylim = c(-0.5, 1)) +
  ggtitle(paste("Detrended Total Renewable Energy Production ACF")) +
  theme_minimal() +
  theme(plot.title = element_text(size = 5))
```

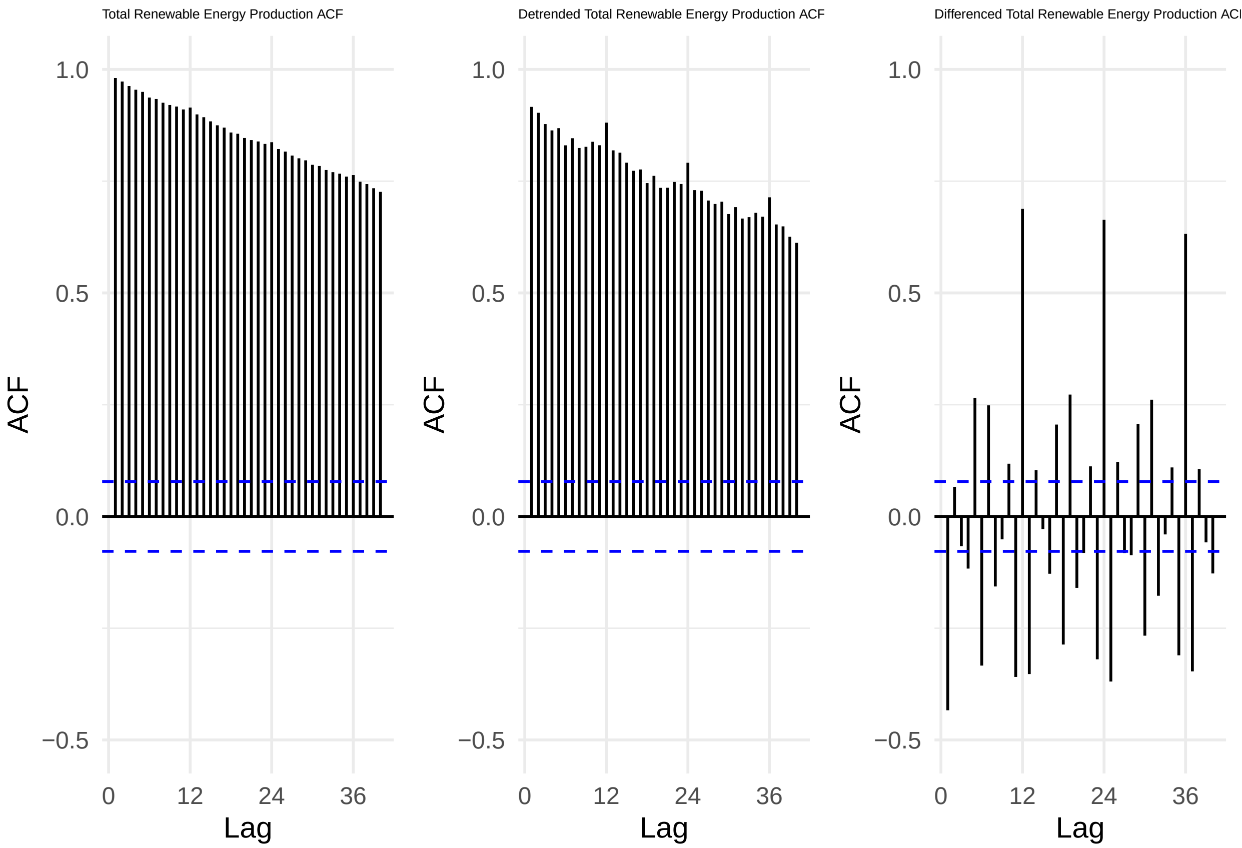
```
## Warning in ggplot2::geom_segment(lineend = "butt", ...): Ignoring unknown
```

```
## parameters: `ylim`
```

```r
energy_diff_acf <- ggAcf(energy_ts_diff, lag.max = 40, ylim = c(-0.5, 1)) +
  ggtitle(paste("Differenced Total Renewable Energy Production ACF")) +
  theme_minimal() +
  theme(plot.title = element_text(size = 5))
```

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

```r
plot_grid(trp_acf, trp_detrended_acf, energy_diff_acf, nrow = 1)
```



Answer: Differencing was more efficient in eliminiating 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 tests. 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
SMKtest <- SeasonalMannKendall(energy_ts)
print("Results for Seasonal Mann Kendall /n")
```

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

```r
print(summary(SMKtest))
```

```
## Score =  13083 , Var(Score) = 201135
```

```
## denominator =  16379.5
## tau = 0.799, 2-sided pvalue =< 2.22e-16
## NULL
```

```r
print("Results for ADF test/n")
```

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

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

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

> Answer: The Seasonal Mann–Kendall test rejects the null hypothesis of no trend (tau = 0.799, p < 2.22e-16), indicating an increasing trend in Total Renewable Energy Production after accounting for seasonality. The Augmented Dickey–Fuller test fails to reject the null hypothesis of a unit root (p = 0.9347), suggesting the series is non-stationary and contains a stochastic trend. These results are consistent with Q3, which indicated that differencing is required to achieve stationarity.

Q6

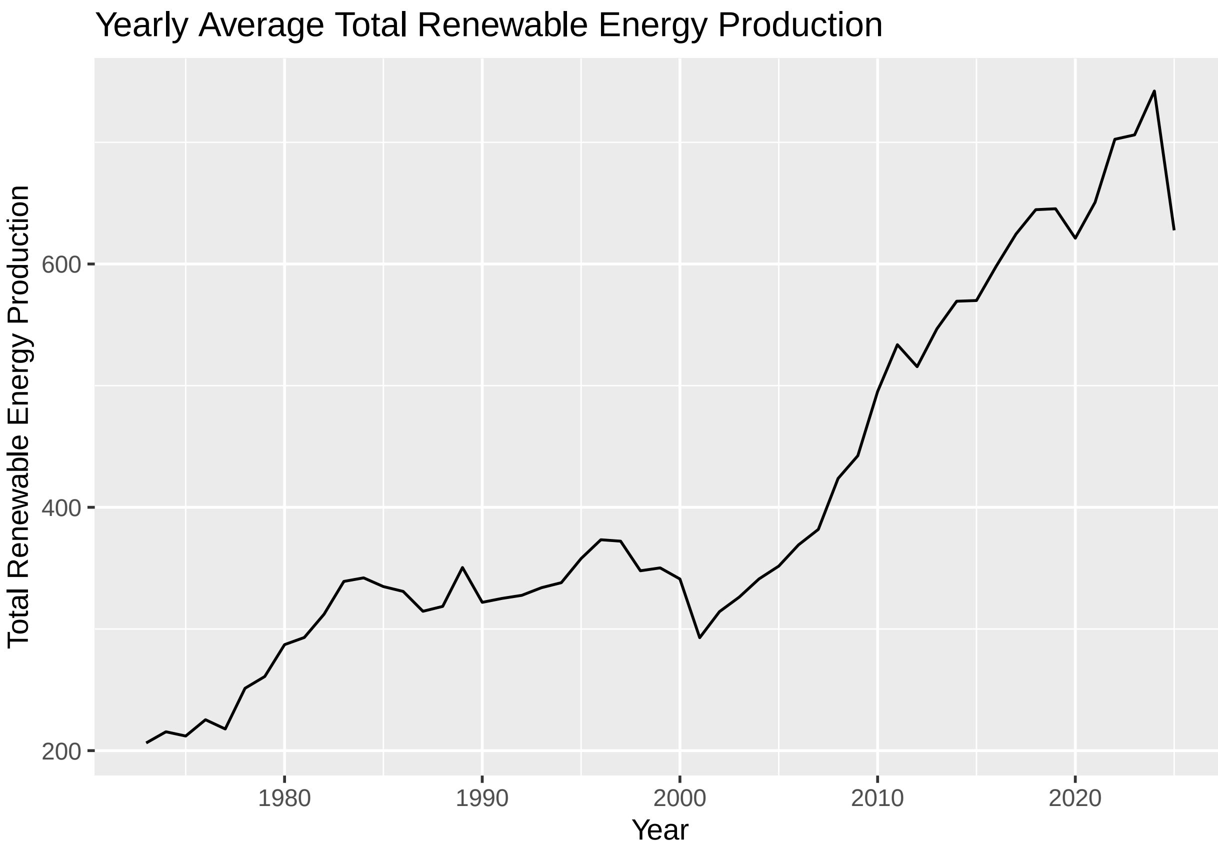
Aggregate the original "Total Renewable Energy Production" series by year. You can use the same procedure we used in class. Store series in a matrix where rows represent months and columns represent years. And then take the columns mean using function colMeans(). Recall the goal is the remove the seasonal 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
energy_subset_vec <- as.numeric(energy_subset$Total_Renewable_Energy_Production)
energy_subset_matrix <- matrix(energy_subset_vec,byrow=FALSE,nrow=12)
```

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

```r
energy_subset_yearly <- colMeans(energy_subset_matrix)
energy_subset_yearly_ts <- ts(energy_subset_yearly, start = 1973, frequency = 1)

autoplot(energy_subset_yearly_ts) +
  ggtitle("Yearly Average Total Renewable Energy Production") +
  xlab("Year") +
  ylab("Total Renewable Energy Production")
```



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?

```
#Mann Kendall Test
MKtest_yearly <- MannKendall(energy_subset_yearly_ts)
print("Results for Mann Kendall /n")
```

```
## [1] "Results for Mann Kendall /n"
```

```
print(summary(MKtest_yearly))
```

```
## Score =  1124 , Var(Score) = 16995.33
## denominator =  1378
## tau = 0.816, 2-sided pvalue =< 2.22e-16
## NULL
```

```
t_yearly <- time(energy_subset_yearly_ts)

#Deterministic trend with Spearman Correlation Test
print("Results from Spearman Correlation")
```

```
## [1] "Results from Spearman Correlation"
```

```
sp_rho=cor(energy_subset_yearly_ts,t_yearly,method="spearman")
print(sp_rho)
```

```
## [1] 0.9234801
```

```r
#with cor.test you can get test statistics
sp_rho=cor.test(energy_subset_yearly_ts,t_yearly,method="spearman")
print(sp_rho)
```

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

```r
print("Results for ADF test/n")
```

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

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

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

> Answer: The Mann–Kendall and Spearman rank tests applied to the yearly aggregated series both indicate a statistically significant increasing trend, consistent with the results obtained for the monthly series in Q5. The ADF test fails to reject the null hypothesis of a unit root, indicating that the yearly series remains non-stationary and contains a stochastic trend. These results agree with Q5.