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

Assignment 5 - Due date 02/28/22

Yared S. Asfaw

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 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).

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_A05_Sp22.Rmd"). Submit this pdf using Sakai.

R packages needed for this assignment are listed below. 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.\

```
#Loading required package
library(forecast)
library(tseries)
library(ggplot2)
library(Kendall)
library(lubridate)
library(tidyverse)
library(dplyr)
```

Decomposing Time Series

Consider the same data you used for A04 from the spreadsheet "Table_10.1_Renewable_Energy_Production_and_Consump
The data comes from the US Energy Information and Administration and corresponds to the January 2021 Monthly Energy Review.

```
#Importing the data set
library(openxlsx)
energy_data <- read.xlsx("./Data/Raw/Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source.xl
startRow=11)
energy_data <- data.frame(energy_data[-1,])
head(energy_data)
```

```
##   Month Wood.Energy.Production Biofuels.Production
## 2 26665                 129.63       Not Available
## 3 26696                117.194       Not Available
```

```
## 4 26724                129.763       Not Available
## 5 26755                125.462       Not Available
## 6 26785                129.624       Not Available
## 7 26816                125.435       Not Available
##   Total.Biomass.Energy.Production Total.Renewable.Energy.Production
## 2                         129.787                           403.981
## 3                         117.338                             360.9
## 4                         129.938                           400.161
## 5                         125.636                            380.47
## 6                         129.834                           392.141
## 7                         125.611                           377.232
##   Hydroelectric.Power.Consumption Geothermal.Energy.Consumption
## 2                         272.703                         1.491
## 3                         242.199                         1.363
## 4                          268.81                         1.412
## 5                         253.185                         1.649
## 6                          260.77                         1.537
## 7                         249.859                         1.763
##   Solar.Energy.Consumption Wind.Energy.Consumption Wood.Energy.Consumption
## 2            Not Available           Not Available                  129.63
## 3            Not Available           Not Available                 117.194
## 4            Not Available           Not Available                 129.763
## 5            Not Available           Not Available                 125.462
## 6            Not Available           Not Available                 129.624
## 7            Not Available           Not Available                 125.435
##   Waste.Energy.Consumption Biofuels.Consumption
## 2                    0.157        Not Available
## 3                    0.144        Not Available
## 4                    0.176        Not Available
## 5                    0.174        Not Available
## 6                     0.21        Not Available
## 7                    0.176        Not Available
##   Total.Biomass.Energy.Consumption Total.Renewable.Energy.Consumption
## 2                          129.787                            403.981
## 3                          117.338                              360.9
## 4                          129.938                            400.161
## 5                          125.636                             380.47
## 6                          129.834                            392.141
## 7                          125.611                            377.232
```

```r
# Formatting the month column to "Date"
start_date <- as.Date("01/01/1973", format = "%m/%d/%Y") # the date format in the dataset is m/d/y
energy_data$Month <- seq(start_date, by= "month", length.out=585)
```

Q1

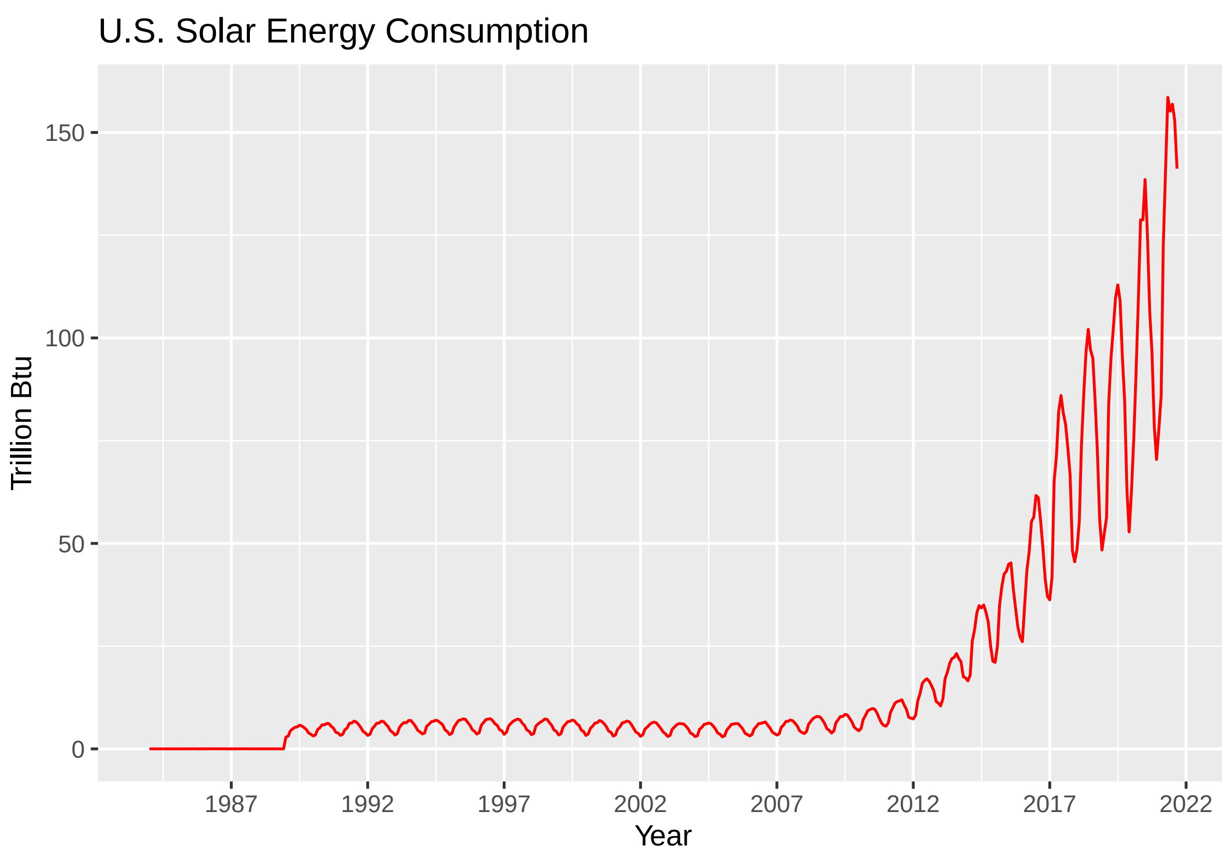
For this assignment you will work only with the following columns: Solar Energy Consumption and Wind Energy Consumption. Create a data frame structure with these two time series only and the Date column. Drop the rows with *Not Available* and convert the columns to numeric. You can use filtering to eliminate the initial rows or convert to numeric and then use the drop_na() function. If you are familiar with pipes for data wrangling, try using it!

```r
  SW_energy_con <- energy_data %>%
  select(Date="Month", SolarEC = "Solar.Energy.Consumption", WindEC = "Wind.Energy.Consumption") %>%
  filter(SolarEC!="Not Available")
SW_energy_con$SolarEC <- as.numeric(SW_energy_con$SolarEC)
SW_energy_con$WindEC <- as.numeric(SW_energy_con$WindEC)
```

Q2

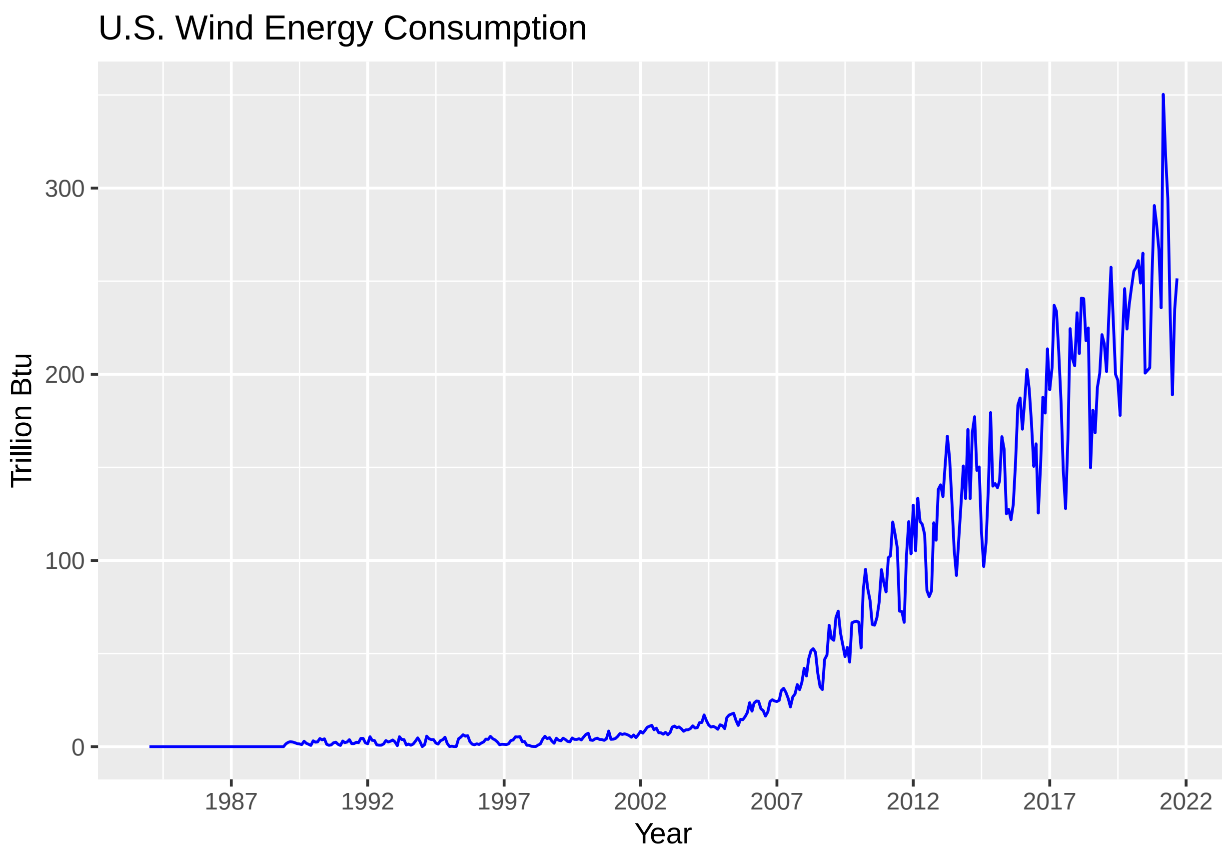
Plot the Solar and Wind energy consumption over time using ggplot. Plot each series on a separate graph. No need to add legend. Add informative names to the y axis using `ylab()`. Explore the function scale_x_date() on ggplot and see if you can change the x axis to improve your plot. Hint: use *scale_x_date(date_breaks = "5 years", date_labels = "%Y")"*)

```r
# Plotting the solar energy consumtpion
ggplot(SW_energy_con, aes(x=as.Date(Date), y=SolarEC))+
  geom_line(color="red")+
  ylab("Trillion Btu") +
  xlab("Year") +
  scale_x_date(date_breaks = "5 years", date_labels = "%Y")+
  ggtitle("U.S. Solar Energy Consumption")
```




```r
# Plotting the wind energy consumtpion
ggplot(SW_energy_con, aes(x=as.Date(Date), y=WindEC))+
  geom_line(color="blue")+
```

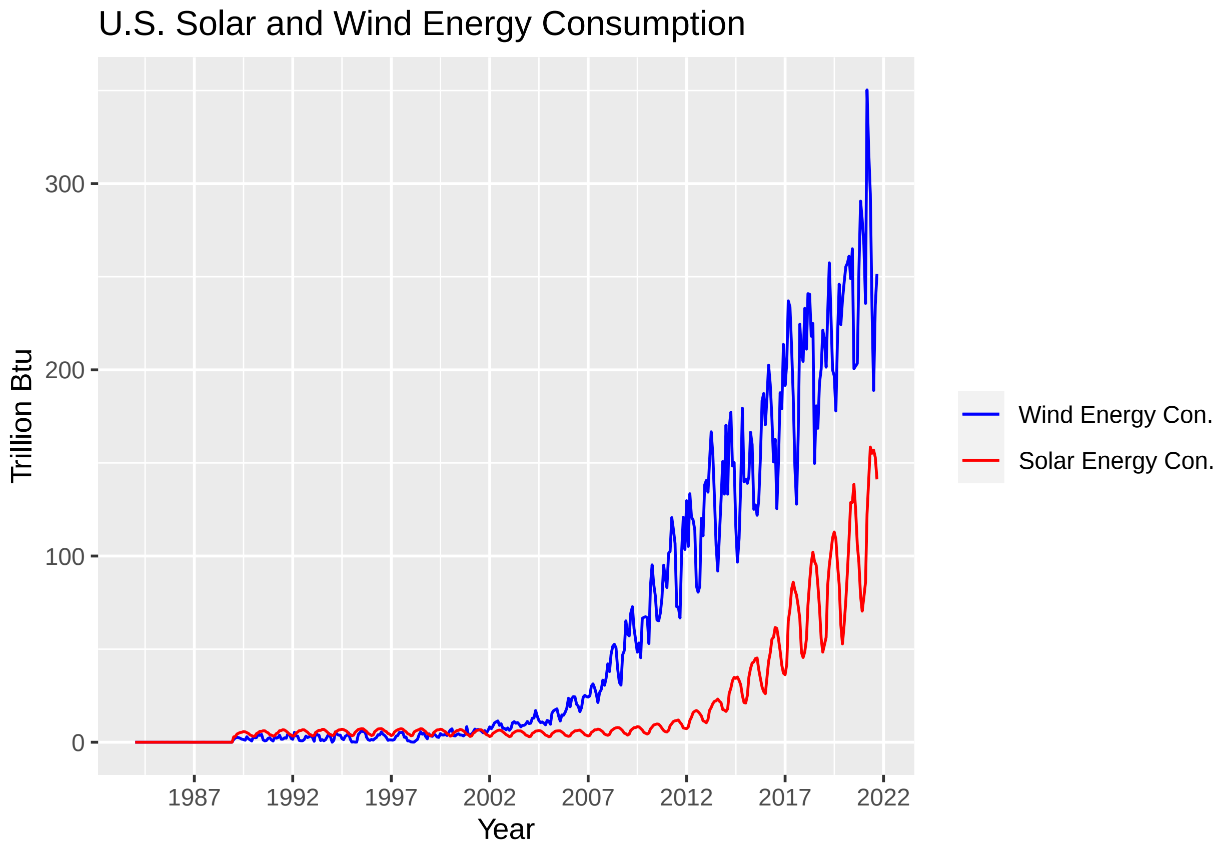
```r
  ylab("Trillion Btu") +
  xlab("Year") +
  scale_x_date(date_breaks = "5 years", date_labels = "%Y")+
  ggtitle("U.S. Wind Energy Consumption")
```



Q3

Now plot both series in the same graph, also using ggplot(). Look at lines 142-149 of the file `05_Lab_OutliersMissingData_Solution` to learn how to manually add a legend to ggplot. Make the solar energy consumption red and wind energy consumption blue. Add informative name to the y axis using `ylab("Energy Consumption)`. And use function scale_x_date() again to improve x axis.

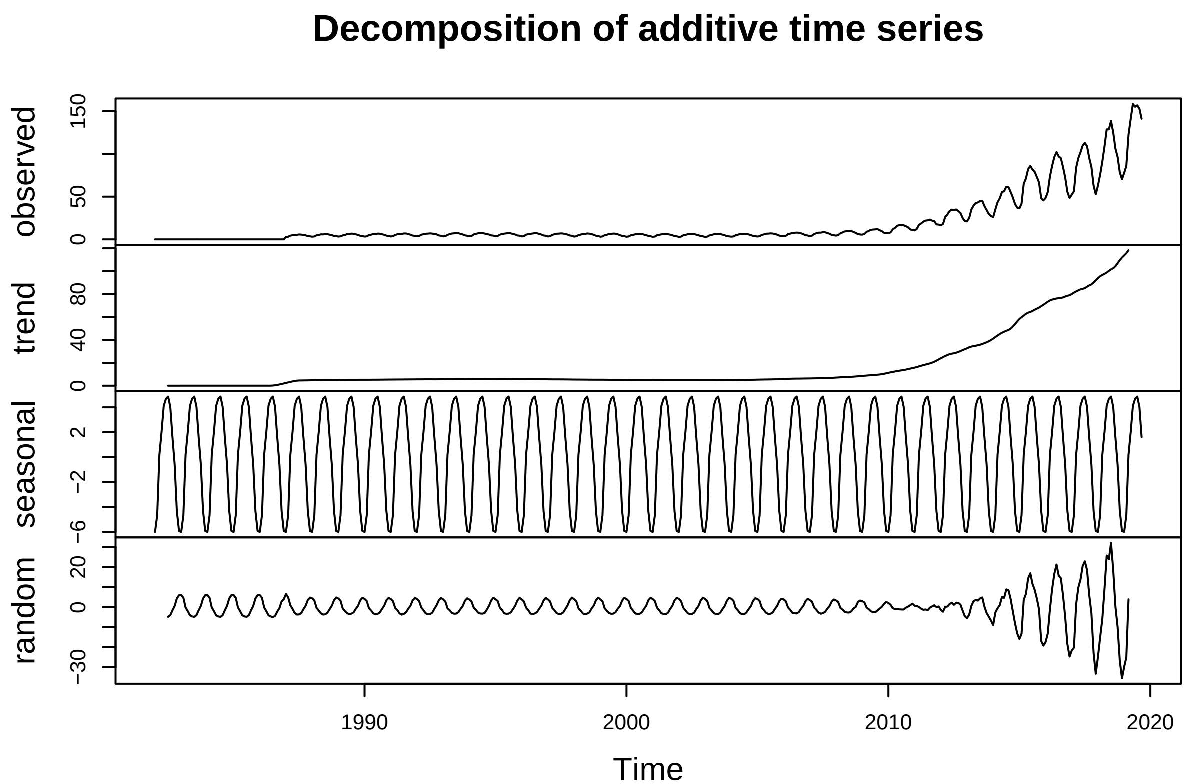
```r
# Plotting both solar and wind energy consumption together
ggplot(SW_energy_con, aes(x=as.Date(Date)))+
  geom_line(aes(y=WindEC, color="WindEC"))+
  geom_line(aes(y=SolarEC, color="SolarEC"))+
  ylab("Trillion Btu") +
  xlab("Year") +
  scale_x_date(date_breaks = "5 years", date_labels = "%Y") +
  ggtitle("U.S. Solar and Wind Energy Consumption") +
  labs(color="") +
  scale_color_manual(values = c("WindEC" = "blue", "SolarEC"="red"),
                     labels=c("Wind Energy Con.", "Solar Energy Con.")) +
  theme(legend.position = "right")
```



Q3

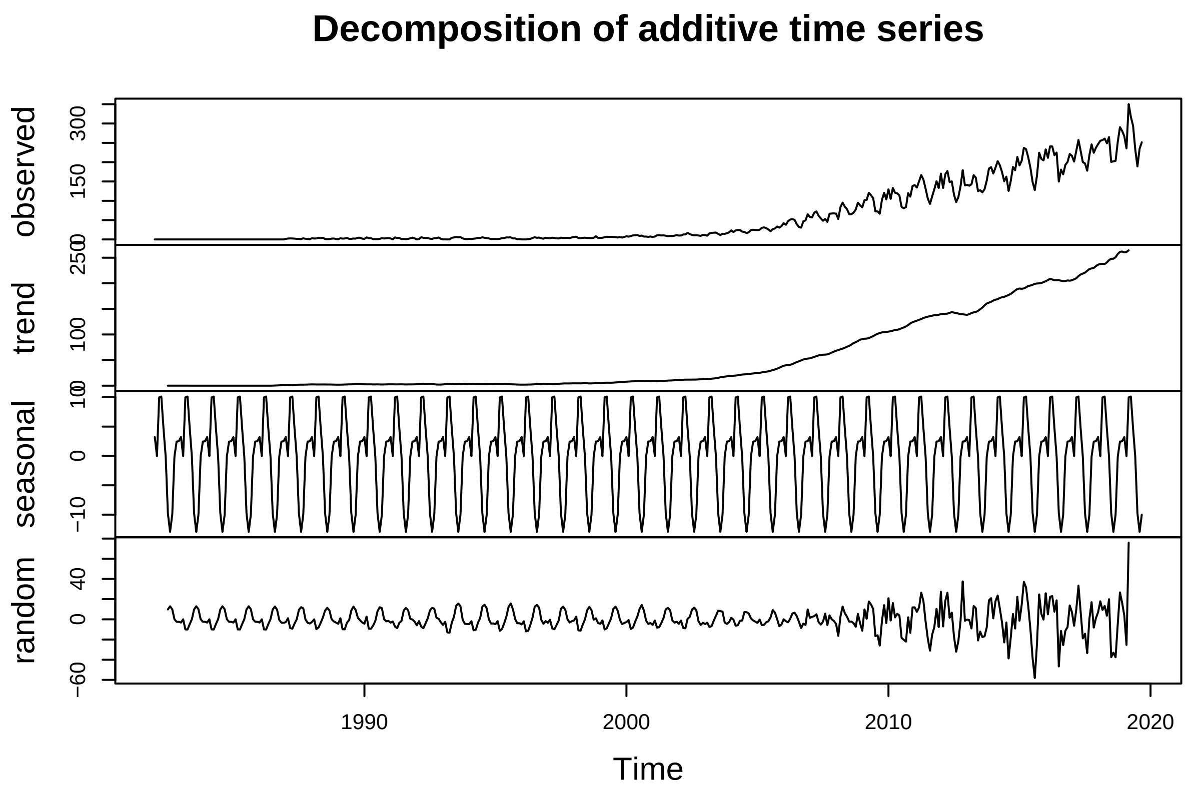
Transform wind and solar series into a time series object and apply the decompose function on them using the additive option, i.e., `decompose(ts_data, type = "additive")`. What can you say about the trend component? What about the random component? Does the random component look random? Or does it appear to still have some seasonality on it?

```
# Transforming the solar data into a time series object and decomposing
ts_solarEC <- ts(SW_energy_con$SolarEC, start = c(1984-01-01), frequency=12)
ts_solarEC_decomposed <- decompose(ts_solarEC, type = "additive")
plot(ts_solarEC_decomposed)
```



The trend component shows that solar energy consumption was almost steadly constant until around the end of 2009 from where it started to dramatically increase overtime. With a similar time period i.e. until the beginning of 2010, the random component shows a seasonal pattern, but after 2010 the randomness of the solar energy consumption become noticable which continued its prononuced random pattern overtime.

```
# Transforming the wind data into a time series object and decomposing
ts_windEC <- ts(SW_energy_con$WindEC, start = c(1984-01-01), frequency=12)
ts_windEC_decomposed <- decompose(ts_windEC, type = "additive")
plot(ts_windEC_decomposed)
```

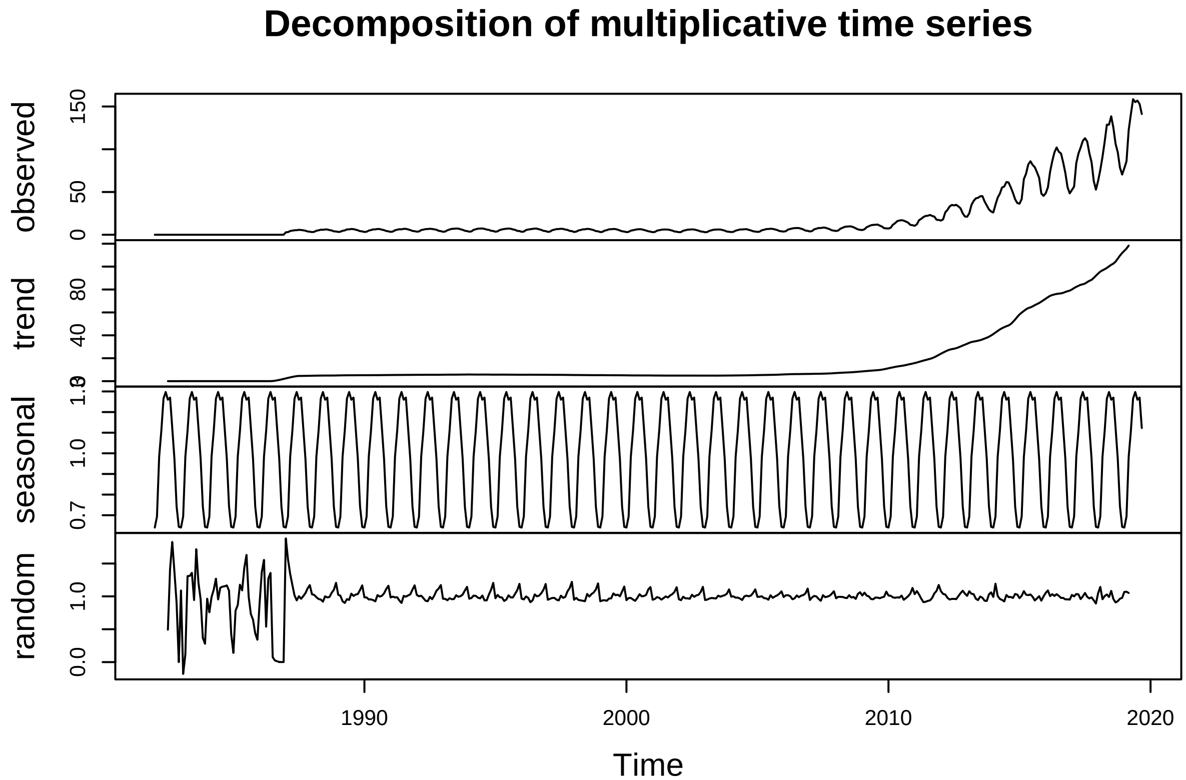


Similar to the trend component in the solar energy consumption, the wind energy consumption also follows the same pattern where it stayed almost steadly constant until the begining of 2000 then after around 2005 it started to dramatically increase overtime. With regard to the random component of the wind energy consumption, until around 2008/09 it showed a seasonal pattern after which the randomness continued to be significantly noticebale overtime.

Q4

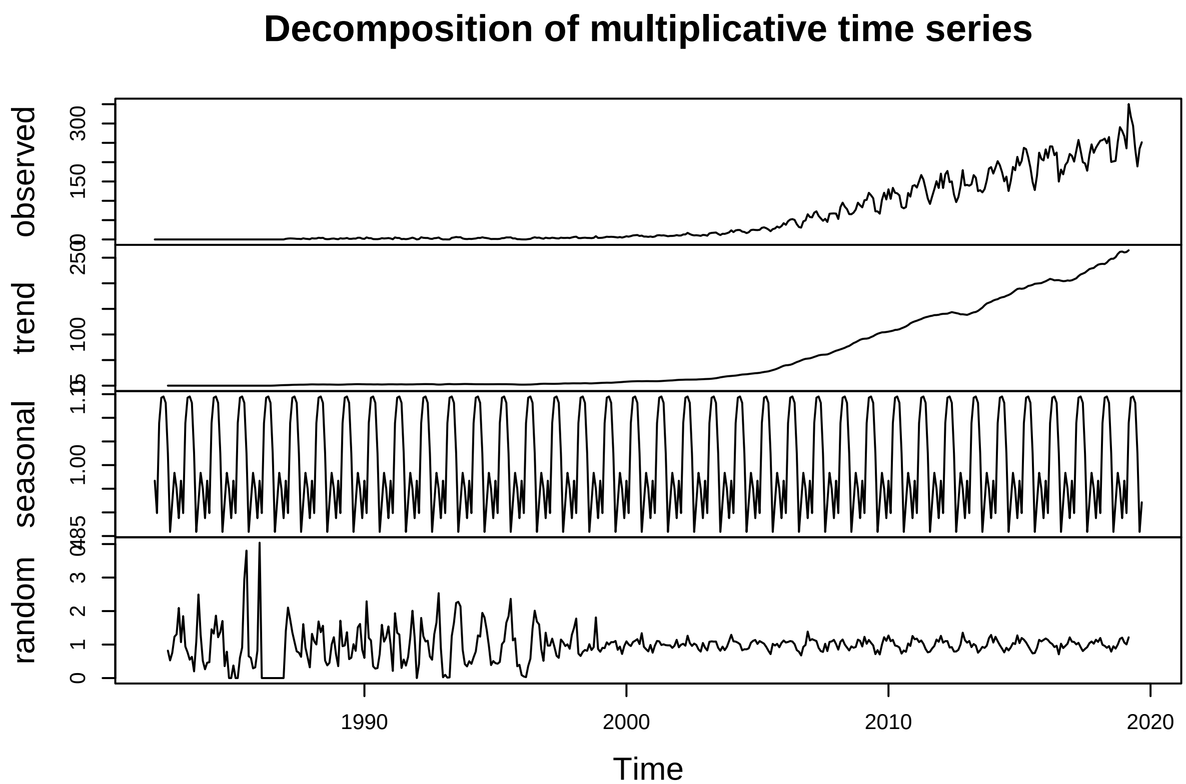
Use the decompose function again but now change the type of the seasonal component from additive to multiplicative. What happened to the random component this time?

```
# Decomposing the solar time series data using multiplicative
ts_solarEC_decomposed <- decompose(ts_solarEC, type = "multiplicative")
plot(ts_solarEC_decomposed)
```



The random component is now significantly noticeable for the period before 1990, after 1990 until around 2005 the random component shows a kind of seasonal pattern, after around 2005 it starts to gain it raondomness slightly overtime but not as significant as the randomness before 1990.

```r
# Decomposing the wind time series data using multiplicative
ts_windEC_decomposed <- decompose(ts_windEC, type = "multiplicative")
plot(ts_windEC_decomposed)
```



In the wind energy consumption time series, the random component is highly noticebale until 2000 where it started to slightly decrease its randomness and from 2010 onwards it it started to show a nearly seasonal pattern.

Q5

When fitting a model to this data, do you think you need all the historical data? Think about the data from 90s and early 20s. Are there any information from those years we might need to forecast the next six months of Solar and/or Wind consumption. Explain your response.

> Answer: In fitting a model, it is good to have as many historical data as possible. However, in this case, using the historical data we have, to fit a model seems misleading (might underestimate the future observations) as the observations in the 90s and early 20s are nearly constant which dramatically changed in trend and randomness later. Thus, in my view I don't think we need all the historical data to fit a model to this data. With regard to forecasting the next six months solar and/or wind energy consumption, for the same reason mentioned above (in the 1990s and early 20s, the consumption of both energy sources was almost constant), the observations in those periods are not relevant for predicting the future trend and/or pattern. Therefore, there is no any information in those years that we need to forecast the next six months solar and/or wind energy consumption.

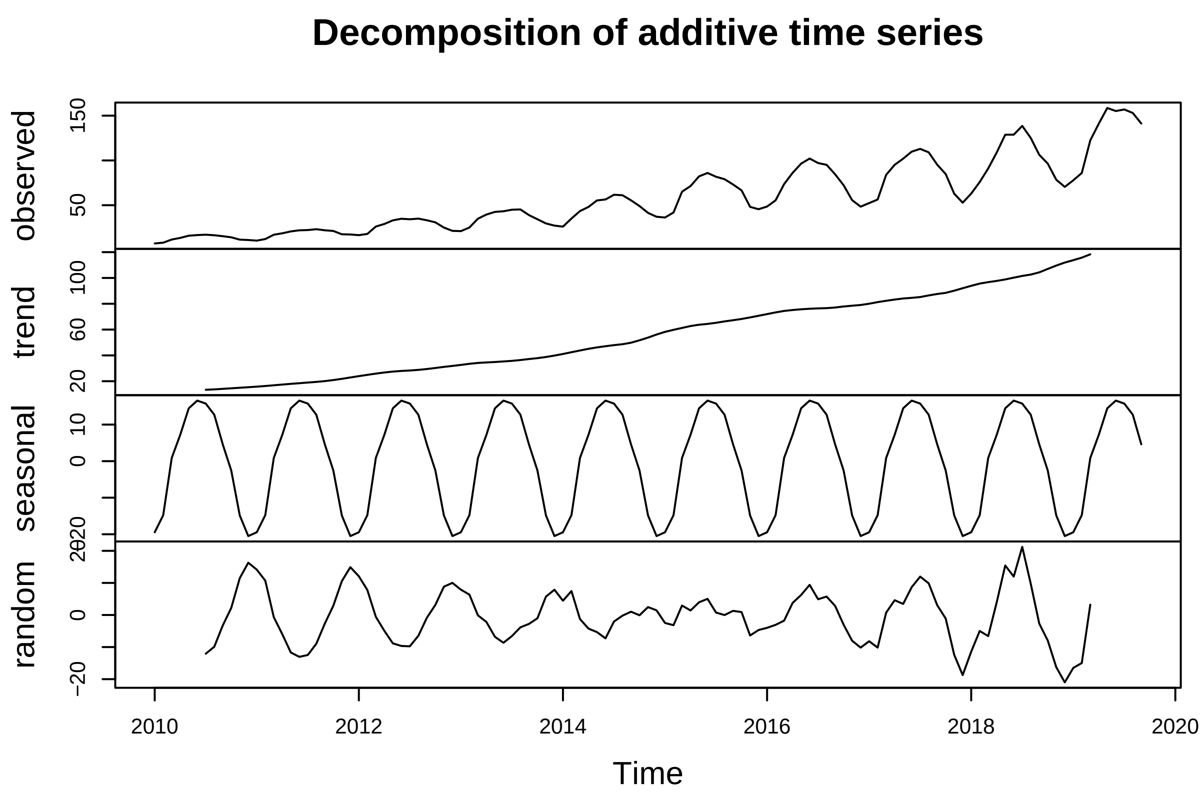
Q6

Create a new time series object where historical data starts on January 2012. Hint: use `filter()` function so that you don't need to point to row numbers, .i.e, `filter(xxxx, year(Date) >= 2012 )`. Apply the decompose function `type=additive` to this new time series. Comment the results. Does the random component look random? Think about our discussion in class about trying to remove the seasonal component and the challenge of trend on the seasonal component.

```
# Creating a new time series for selected period and decomposing using additive type
SW_energy_con_recent <- SW_energy_con %>%
  filter(year(Date)>=2012)
head(SW_energy_con_recent)
```

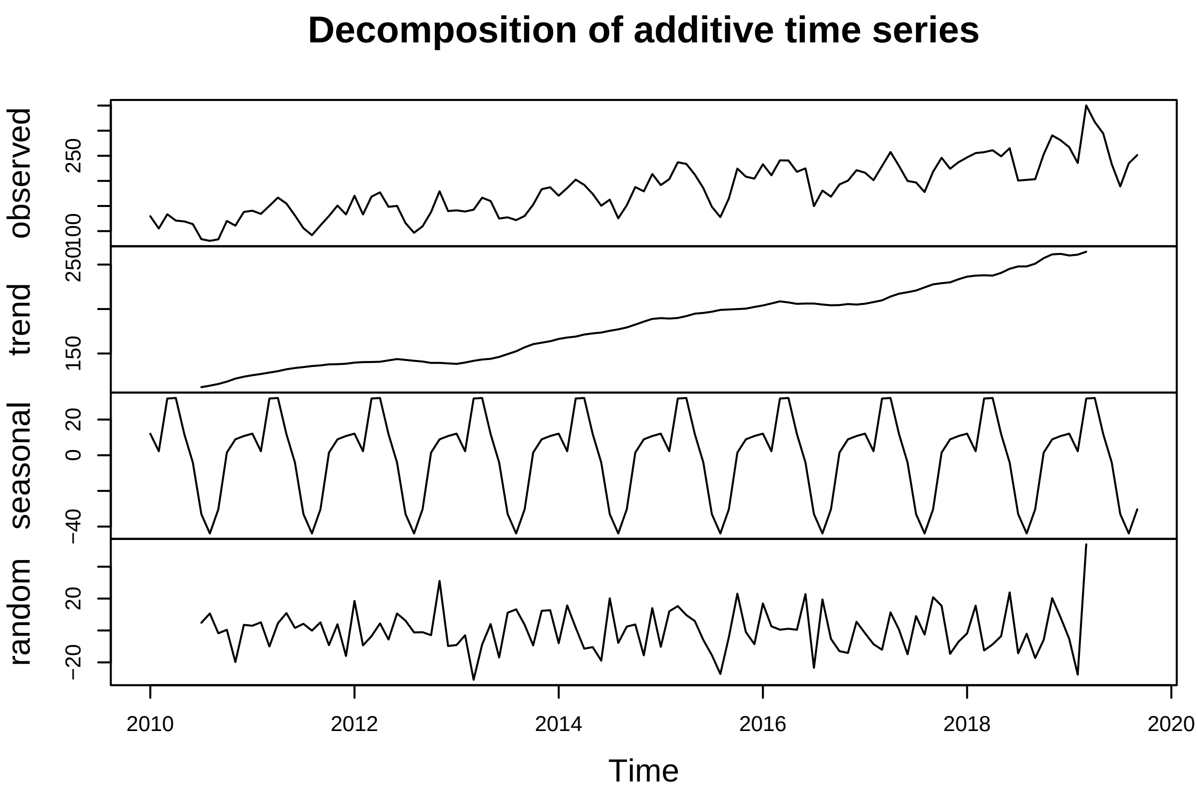
```
##          Date SolarEC  WindEC
## 1 2012-01-01   7.288 129.726
## 2 2012-02-01   8.165 105.171
## 3 2012-03-01  11.678 133.476
## 4 2012-04-01  13.478 120.941
## 5 2012-05-01  15.933 119.336
## 6 2012-06-01  16.651 113.928
```

```
# Solar energy consumption
ts1_solarEC <- ts(SW_energy_con_recent$SolarEC, start = c(2012-01-01), frequency = 12)
ts1_solarEC_decomposed <- decompose(ts1_solarEC, type = "additive")
plot(ts1_solarEC_decomposed)
```



The decomposition of additive time series of solar energy consumption for the observations starting from 2012 show a clear dramatically increasing trend of the solar energy consumption.The random component on the other hand shows a kind of seasonal pattern at the beginning until 2013 and then turns to a clear random behavior starting from the end of 2013.

```
# Wind energy consumption
ts1_windEC <- ts(SW_energy_con_recent$WindEC, start = c(2012-01-01), frequency = 12)
ts1_windEC_decomposed <- decompose(ts1_windEC, type = "additive")
plot(ts1_windEC_decomposed)
```



The decomposition of additive time series of wind energy consumption for the observations starting from 2012 show an increasing trend of the wind energy consumption overtime. Similarly, the random component shows a completely clear random behavior from start to end.

Answer: Therefore, for the solar energy consumption, the random component now clearly exhibits randomness starting from around the end of 2013. On the other hand, the random component of the wind energy concumption showed a completely random pattern from start to end.