ENV 790.30 - Time Series Analysis for Energy Data | Spring 2023

Assignment 5 - Due date 02/27/23

Maggie O'Shea

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_A05_Sp23.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(forecast)
library(tseries)
library(ggplot2)
library(Kendall)
library(lubridate)
library(tidyverse)  #load this package so yon clean the data frame using pipes
```

Decomposing Time Series

Consider the same data you used for A04 from the spreadsheet "Table_10.1_Renewable_Energy_Production_and_Consump

```
#Importing data set - using xlsx package

energy_data<- read_excel("./Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source.xlsx"
#Note: Package xlsx does not work because of the java on my computer. I used readxl, which requires sli


#Now let's extract the column names from row 11 only
read_col_names <- read_xlsx(path="Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source
colnames(energy_data) <- read_col_names
```

```r
#head(energy_data)

nobs=nrow(energy_data)
nvar=ncol(energy_data)
```

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
energy_data_clean <- energy_data%>%
  filter(`Solar Energy Consumption` != "Not Available",
         `Wind Energy Consumption` != "Not Available")%>%
  mutate("Solar Energy Consumption" = as.numeric(`Solar Energy Consumption`),
         "Wind Energy Consumption" = as.numeric(`Wind Energy Consumption`))%>%
  select(Month, `Solar Energy Consumption`, `Wind Energy Consumption`)
```

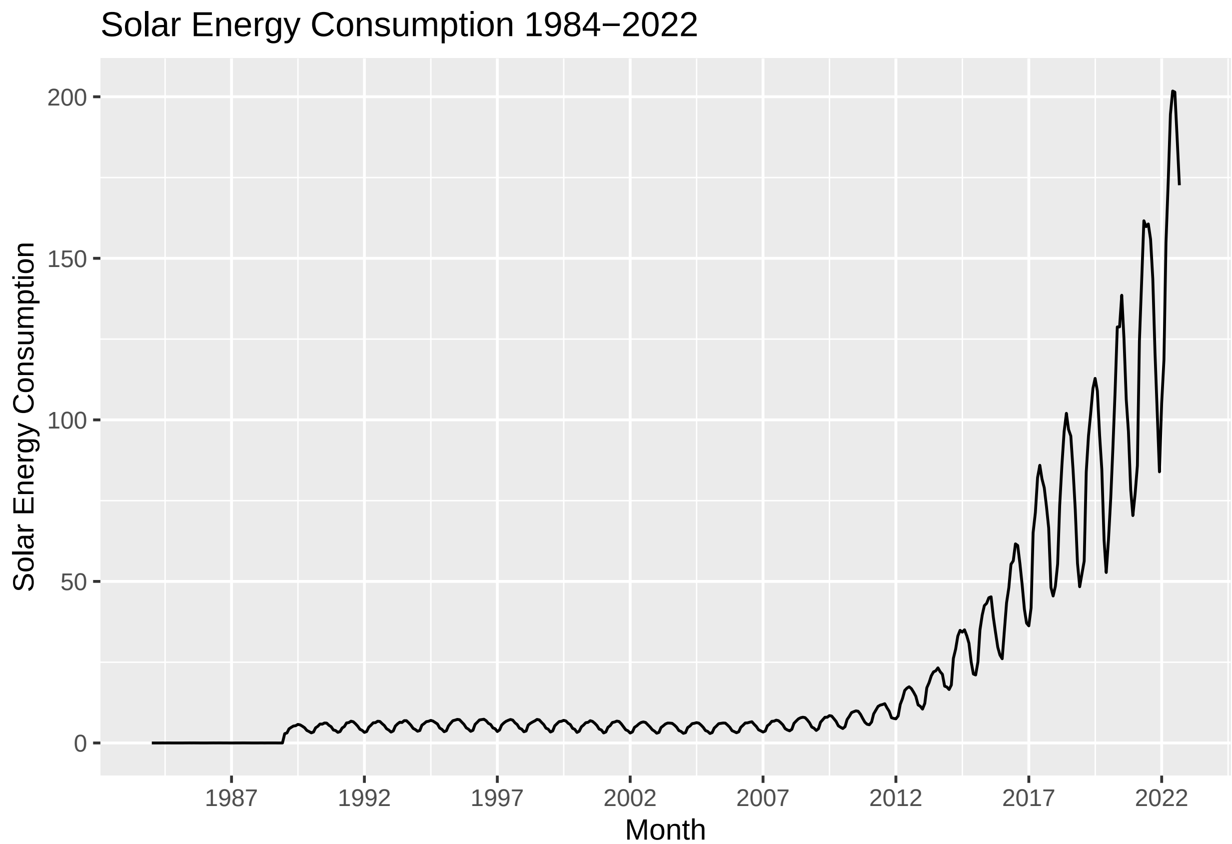
Q2

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")")*

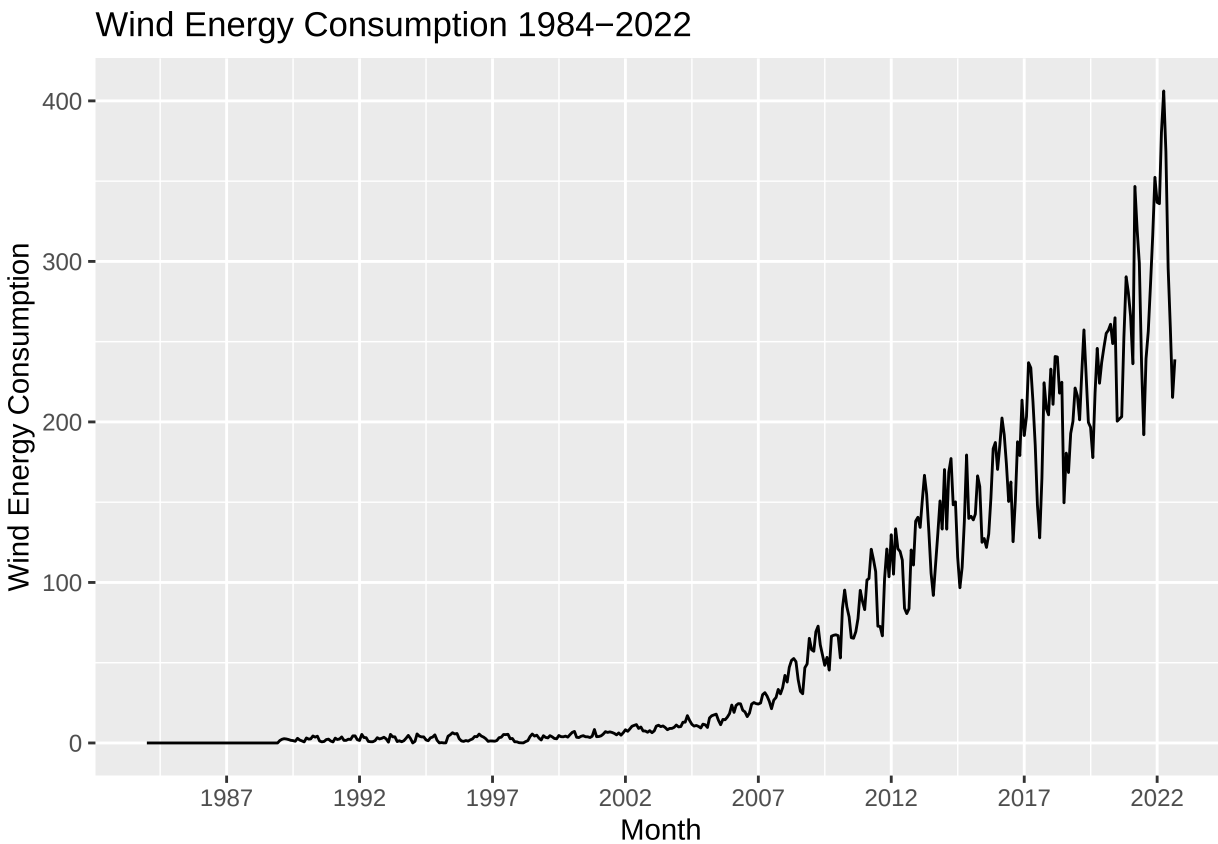
```r
#In order to use scale_x_date() I had to convert the month column to a date
energy_data_clean$Month <- as.Date(energy_data_clean$Month, "%Y-%M-%D")
```

```
## Warning in as.POSIXlt.POSIXct(x, tz = tz): unknown timezone '%Y-%M-%D'
```

```r
ggplot(energy_data_clean, aes(x=Month))+
  geom_line(aes(y=`Solar Energy Consumption`))+
  labs(title = "Solar Energy Consumption 1984-2022", ylab = "Solar Energy Consumption")+
  scale_x_date(date_breaks = '5 years', date_labels = '%Y')
```



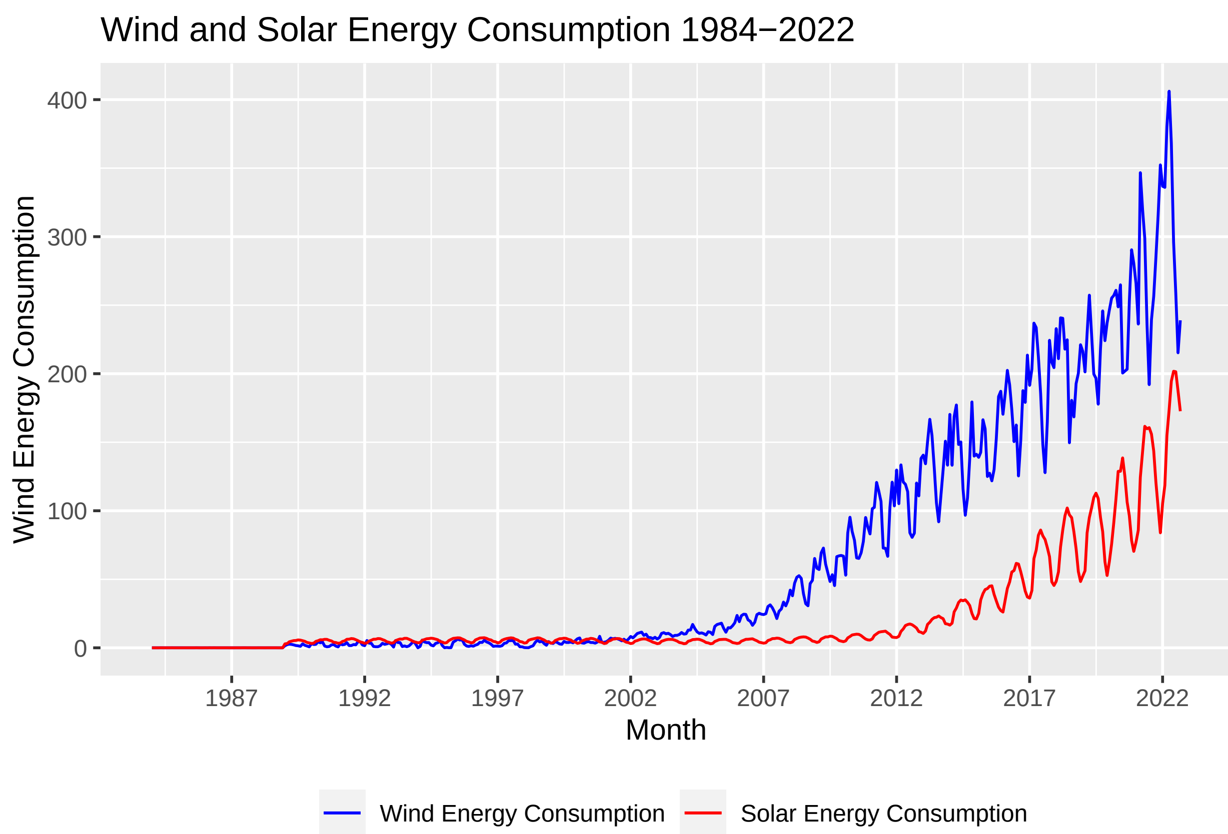

```
ggplot(energy_data_clean, aes(x=Month))+
  geom_line(aes(y=`Wind Energy Consumption`))+
  labs(title = "Wind Energy Consumption 1984-2022", ylab = "Wind Energy Consumption")+
  scale_x_date(date_breaks = '5 years', date_labels = '%Y')
```



Q3

Now plot both series in the same graph, also using ggplot(). Look at lines 141-148 of the file `M4_OutliersMissingData_Part2_Complete.Rmd` 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.

```
ggplot(energy_data_clean, aes(x=Month))+
  geom_line(aes(y=`Wind Energy Consumption`, color = "Wind Energy Consumption"))+
  geom_line(aes(y=`Solar Energy Consumption`, color = "Solar Energy Consumption"))+
  labs(color="") +
  scale_color_manual(values = c("Wind Energy Consumption" = "blue", "Solar Energy Consumption" = "red")
  theme(legend.position = "bottom") +
  labs(title = "Wind and Solar Energy Consumption 1984-2022", ylab = "Energy Consumption")+
  scale_x_date(date_breaks = '5 years', date_labels = '%Y')
```

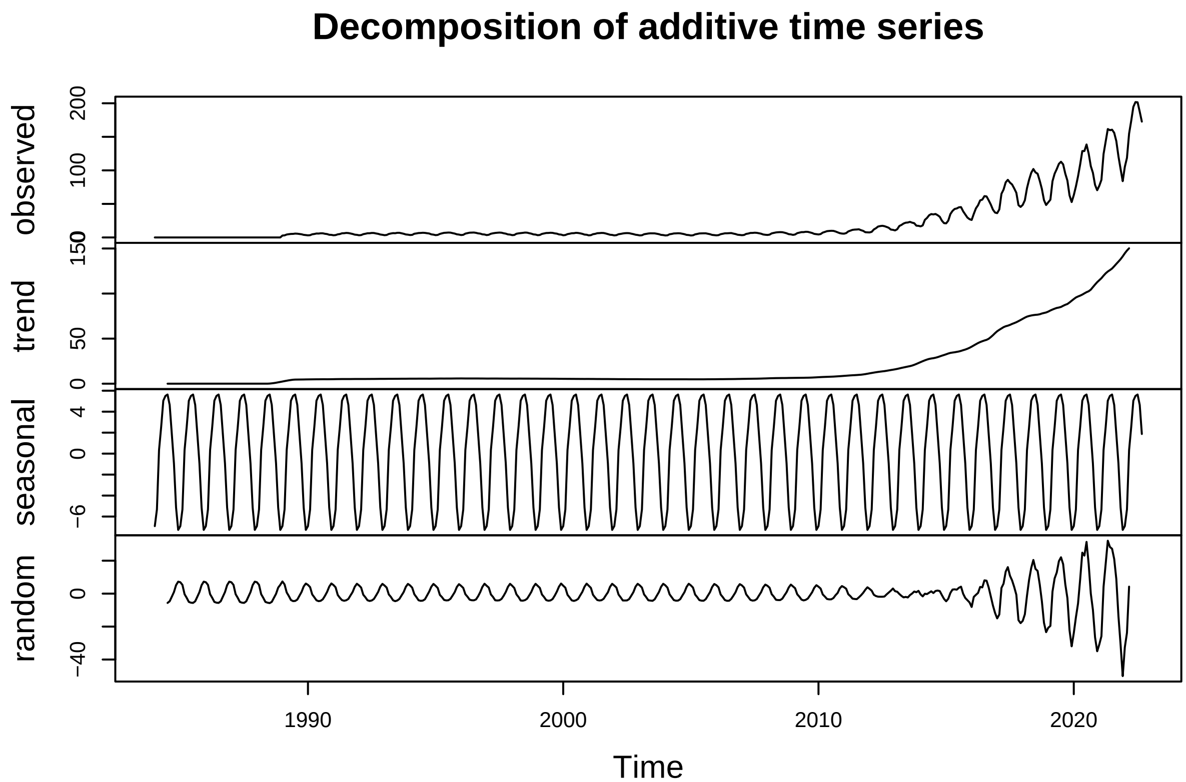


Q3

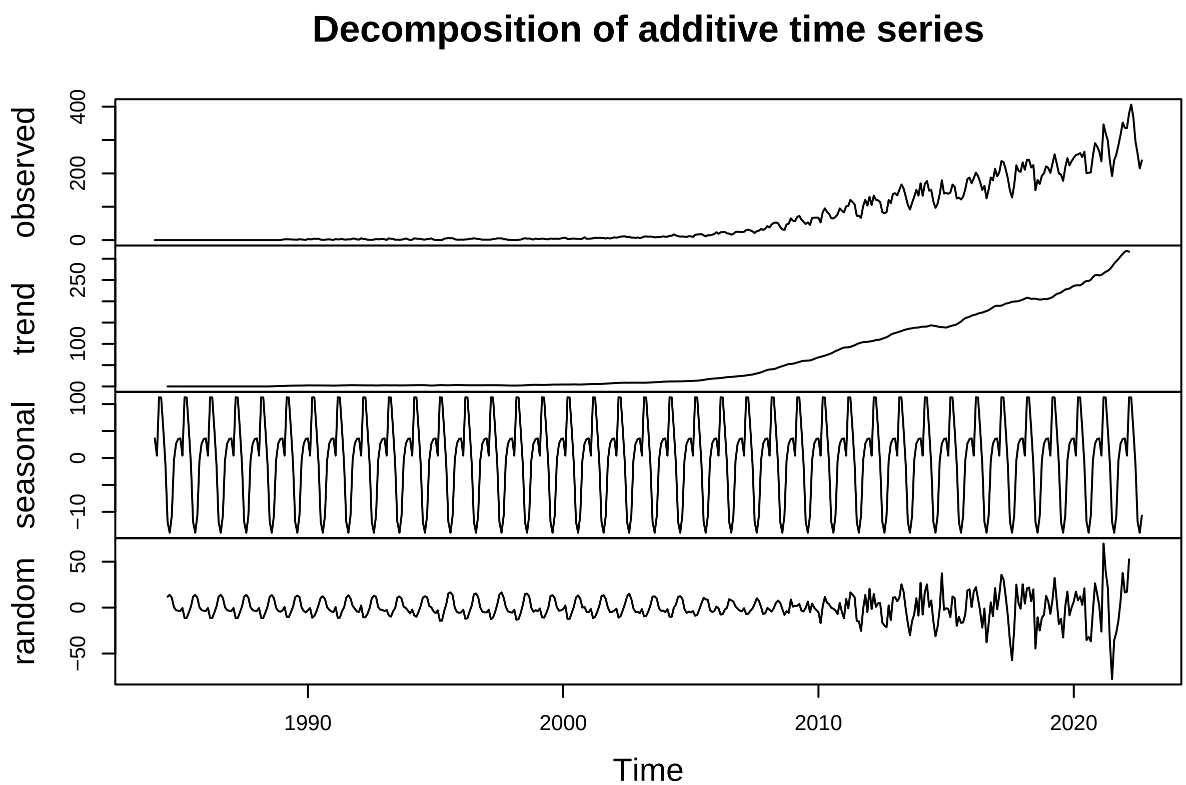
Transform wind and solar series into a time series object and apply the decompose function on them using the additive option, i.e., `decompose(ts_data, type = "additive")`. What can you say about the trend component? What about the random component? Does the random component look random? Or does it appear to still have some seasonality on it? **Add to answer**: - If you do an additive model on a dataset that is multiplicative it will look like there is still a seasonality to it

```
ts_solar <- ts(energy_data_clean[,2], frequency = 12, start = c(1984, 1))
ts_wind <- ts(energy_data_clean[,3], frequency = 12, start = c(1984, 1))

plot(decompose(ts_solar, type = "additive"))
```




```r
plot(decompose(ts_wind, type = "additive"))
```



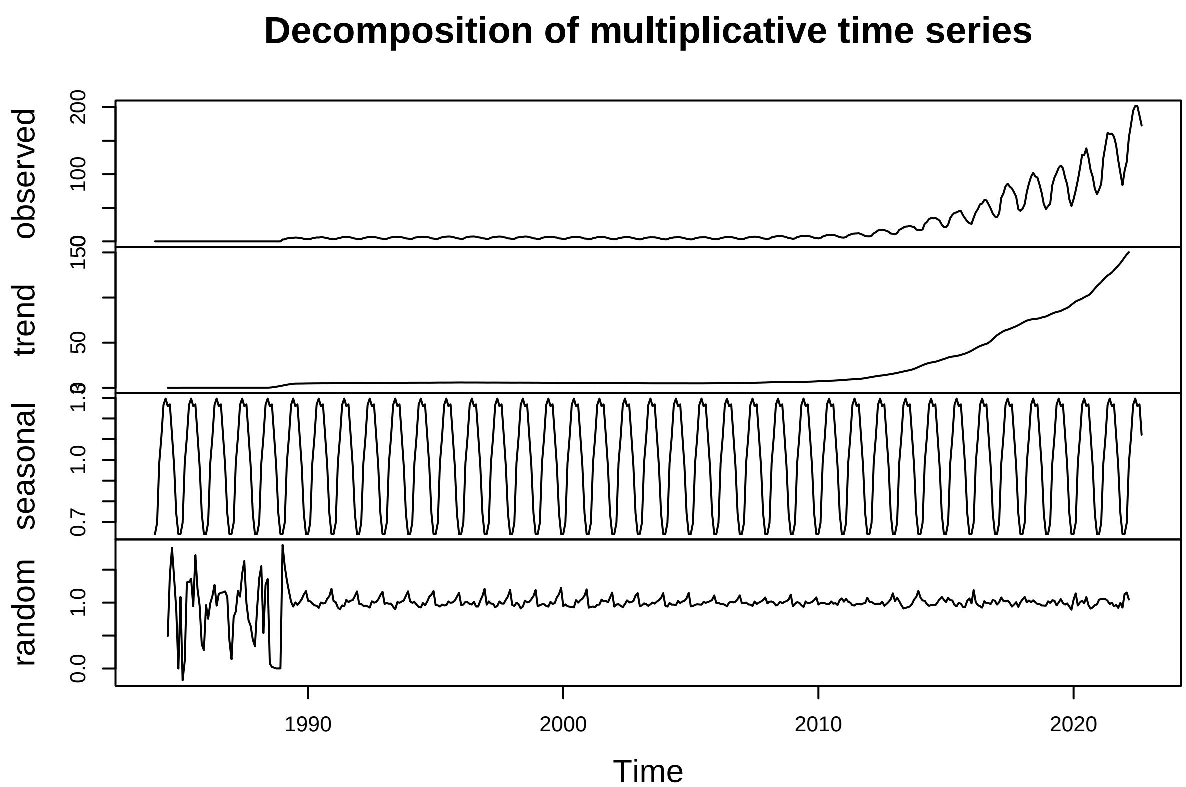
Q3 Answer:

Answer: The trend component for solar appears to be an exponential increase. In addition, the random component in the solar model does appear to still have some seasonality in it and have trend in it, with the magnitude or variation in values widening around 2015-2022. This suggests the multiplicative model would be a better fit for the solar data. The trend component for wind also appears to be an exponential increase. There also seems to still be a seasonality in the random component for wind, and the same change in magnitude occurs near 2010/2012 in the random component. Multiplicative model for wind would likely be a better fit as well.

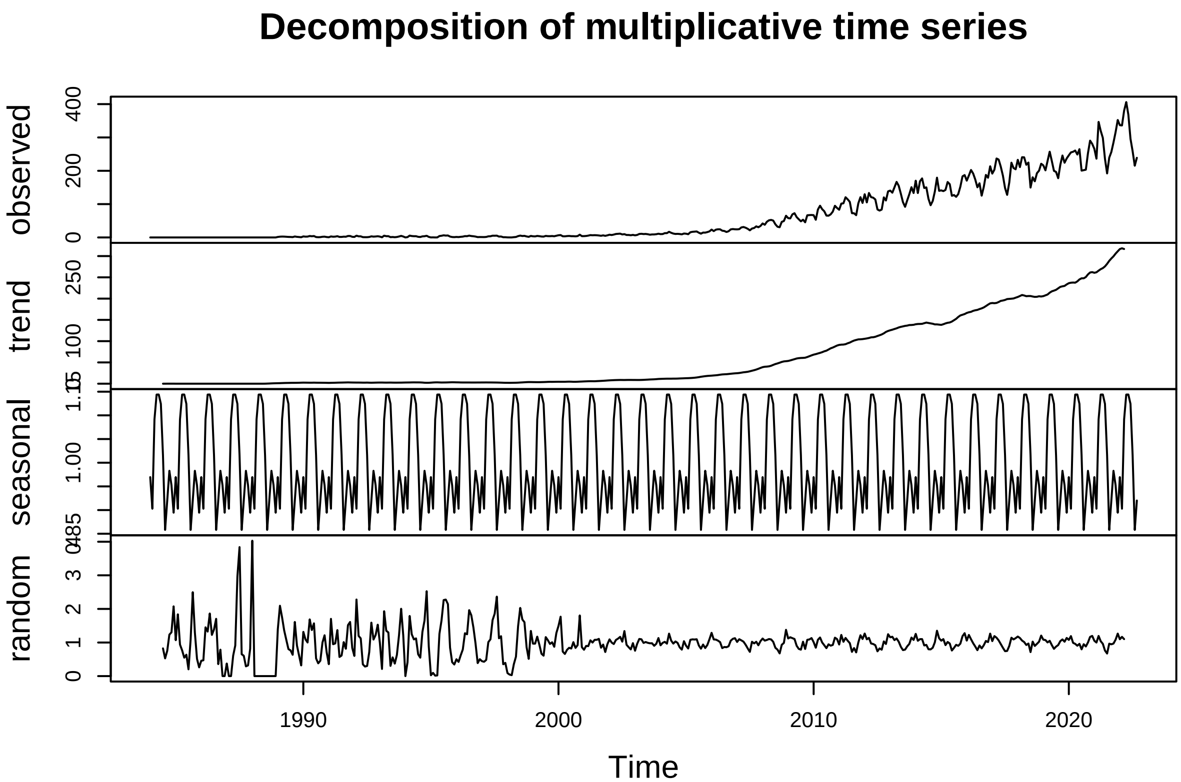
Q4

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?

```
plot(decompose(ts_solar, type = "multiplicative"))
```




```
plot(decompose(ts_wind, type = "multiplicative"))
```



Q4 Answer:

Answer:The decomposed solar time series random component shows the opposite pattern as it did when it was additive. When changing to multiplicative, the random component for the solar time series shows more variation at the beginning and then becomes much smaller by 1990. (The opposite was true in the additive model where the variation in values got much bigger at the end of the series.) In addition, these random component values hover around 1.0 rather than 0, like the additive random component. Finally, the range of random component values is greater in the additive plot which had values from 20 to negative 20, whereas the multiplicative random component ranged from 0 to 2.

The decomposed wind series showed a similar change when changing from additive to multiplicatve. There was more variation in the values at the beginning and less as time increases. Again, these values were hovering around 1 rather than 0.0, and the differences between the values were not as great. In the additive plot the random component moved from 50 to negative 50, whereas with multiplicative the numbers range from 0 to 4.

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.

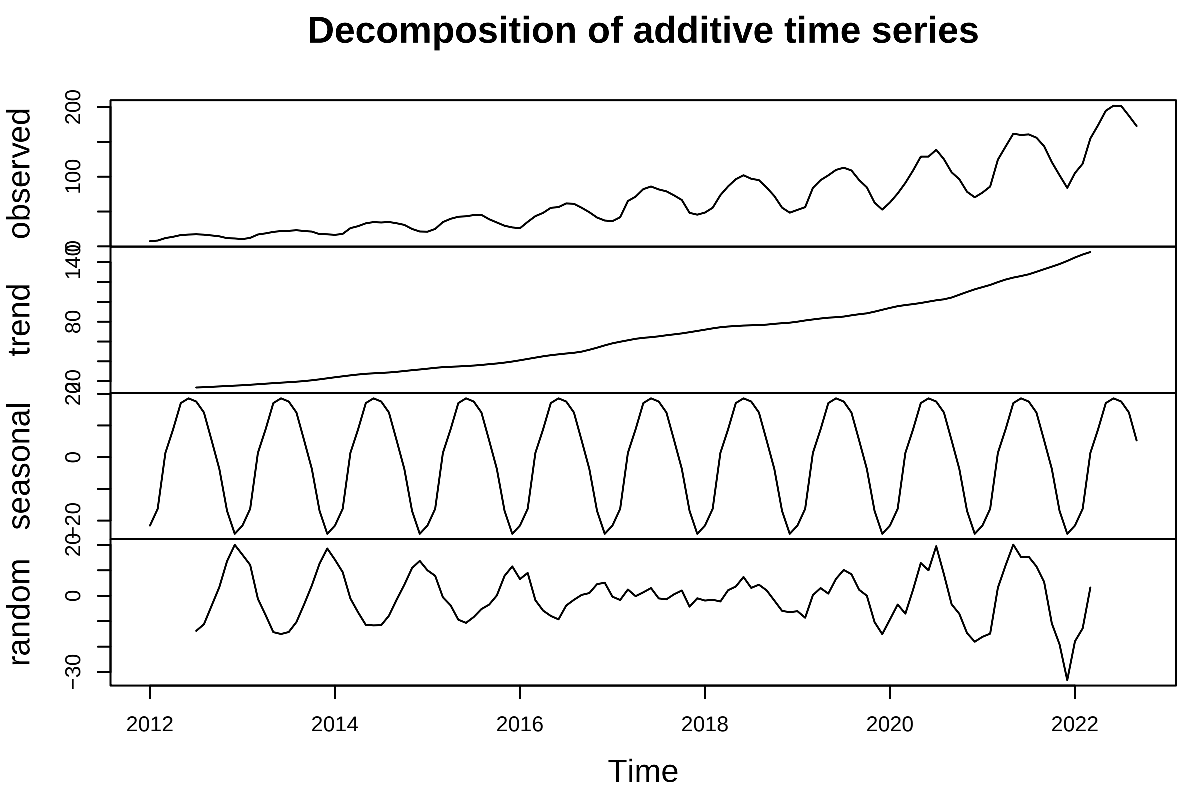
Answer: In examining the plots, it looks like the trend of increase occurred just after 2010. The patterns in the time series seem distinct before and after this point, which makes sense conceptually as well. Since 2010 there likely has been different drivers of change in the renewables industry compared to the 1990s and thus we only want to examine data that is showing relevant patterns that might represent future trends as well, which would be the 2010 and onward data. Solar and wind energy was not widely used before 2010, as compared to today and our understanding of their importance (and thus programs policies that support renewable development) has grown immensely. For this reason, the solar and wind industries of today is quite different from that of the 1990s and therefore the 1990/2000 data may not be necessary to include.

Q6

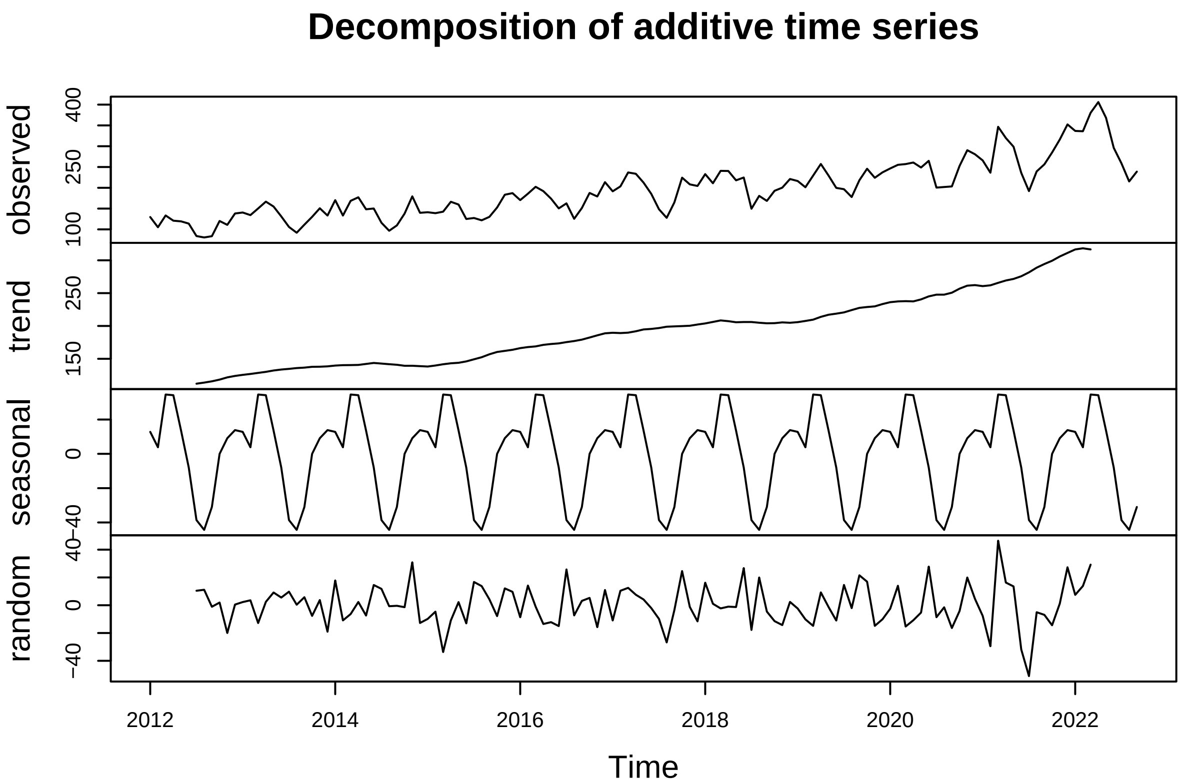
Create a new time series object where historical data starts on January 2012. Hint: use `filter()` function so that you don't need to point to row numbers, .i.e, `filter(xxxx, year(Date) >= 2012 )`. Apply the decompose function `type=additive` to this new time series. Comment the results. Does the random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

```
energy_data_clean2012 <- energy_data_clean%>%
  filter(year(Month)>=2012)
ts_solar2012 <- ts(energy_data_clean2012[,2], frequency=12, start = c(2012, 1, 1))
ts_wind2012 <- ts(energy_data_clean2012[,3], frequency=12, start = c(2012, 1, 1))

plot(decompose(ts_solar2012, type="additive"))
```




```
#plot(decompose(ts_solar2012, type="multiplicative"))
plot(decompose(ts_wind2012, type="additive"))
```



Q6 Answer:

Answer: The results of the 2012-2022 decomposed solar time series with the additive model shows a clear upward trend component and a seasonality plot ranging from -20 to 20. The random component roughly hovers around zero with some spikes (especially in 2022). However, it still shows seasonality. In an additive model the magnitude of seasonality does not change in relation to time, however in a multiplicative model the magnitude of the seasonality depends on the magnitude of the data. Because this random component shows some remaining seasonality, it is clear that this series requires that the magnitude of the seasonality can vary depending on the magnitude of the data. This suggests that a multiplicative model may be better suited.

In contrast, the wind decomposition with the additive model does not show the same challenges. The random component does seem to have a mean of zero, although there are some spikes especially just before 2022. There is not an apparent seasonal appearance to the random component, suggesting that additive could be the right model type.