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

Assignment 5 - Due date 02/27/23

Katherine Burley

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(xlsx)
library(readxl)
library(forecast)
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```
library(tseries)
library(ggplot2)
library(Kendall)
library(lubridate)
```

```
##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
library(tidyverse)  #load this package so you can clean the data frame using pipes
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2
## --

## v tibble  3.1.8     v dplyr   1.1.0
## v tidyr   1.3.0     v stringr 1.5.0
```

```
## v readr   2.1.3     v forcats 1.0.0
## v purrr   1.0.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x lubridate::as.difftime() masks base::as.difftime()
## x lubridate::date()        masks base::date()
## x dplyr::filter()          masks stats::filter()
## x lubridate::intersect()   masks base::intersect()
## x dplyr::lag()             masks stats::lag()
## x lubridate::setdiff()     masks base::setdiff()
## x lubridate::union()       masks base::union()
```

```
# Set option for text to wrap in PDF output
knitr::opts_chunk$set(tidy.opts = list(width.cutoff = 60), tidy = TRUE)
```

Decomposing Time Series

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

```
# Importing data set - using xlsx package
energy_data <- as.data.frame(read_excel(path = "../Data/Table_10.1_Renewable_Energy_Production_and_Cons
    skip = 12, sheet = "Monthly Data", col_names = FALSE))
```

```
## New names:
## * `` -> `...1`
## * `` -> `...2`
## * `` -> `...3`
## * `` -> `...4`
## * `` -> `...5`
## * `` -> `...6`
## * `` -> `...7`
## * `` -> `...8`
## * `` -> `...9`
## * `` -> `...10`
## * `` -> `...11`
## * `` -> `...12`
## * `` -> `...13`
## * `` -> `...14`
```

```
# Extract the column names from row 11
read_col_names <- read_excel(path = "../Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_
    skip = 10, n_max = 1, sheet = "Monthly Data", col_names = FALSE)
```

```
## New names:
## * `` -> `...1`
## * `` -> `...2`
## * `` -> `...3`
## * `` -> `...4`
## * `` -> `...5`
## * `` -> `...6`
## * `` -> `...7`
## * `` -> `...8`
## * `` -> `...9`
## * `` -> `...10`
## * `` -> `...11`
## * `` -> `...12`
## * `` -> `...13`
```

```
## * `` -> `...14`
```

```r
colnames(energy_data) <- read_col_names
head(energy_data)
```

```
##        Month Wood Energy Production Biofuels Production
## 1 1973-01-01                129.630       Not Available
## 2 1973-02-01                117.194       Not Available
## 3 1973-03-01                129.763       Not Available
## 4 1973-04-01                125.462       Not Available
## 5 1973-05-01                129.624       Not Available
## 6 1973-06-01                125.435       Not Available
##   Total Biomass Energy Production Total Renewable Energy Production
## 1                         129.787                           403.981
## 2                         117.338                           360.900
## 3                         129.938                           400.161
## 4                         125.636                           380.470
## 5                         129.834                           392.141
## 6                         125.611                           377.232
##   Hydroelectric Power Consumption Geothermal Energy Consumption
## 1                         272.703                         1.491
## 2                         242.199                         1.363
## 3                         268.810                         1.412
## 4                         253.185                         1.649
## 5                         260.770                         1.537
## 6                         249.859                         1.763
##   Solar Energy Consumption Wind Energy Consumption Wood Energy Consumption
## 1            Not Available           Not Available                 129.630
## 2            Not Available           Not Available                 117.194
## 3            Not Available           Not Available                 129.763
## 4            Not Available           Not Available                 125.462
## 5            Not Available           Not Available                 129.624
## 6            Not Available           Not Available                 125.435
##   Waste Energy Consumption Biofuels Consumption
## 1                    0.157        Not Available
## 2                    0.144        Not Available
## 3                    0.176        Not Available
## 4                    0.174        Not Available
## 5                    0.210        Not Available
## 6                    0.176        Not Available
##   Total Biomass Energy Consumption Total Renewable Energy Consumption
## 1                          129.787                            403.981
## 2                          117.338                            360.900
## 3                          129.938                            400.161
## 4                          125.636                            380.470
## 5                          129.834                            392.141
## 6                          125.611                            377.232
```

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

energy_data$Date <- ymd(energy_data$Month)
```

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!

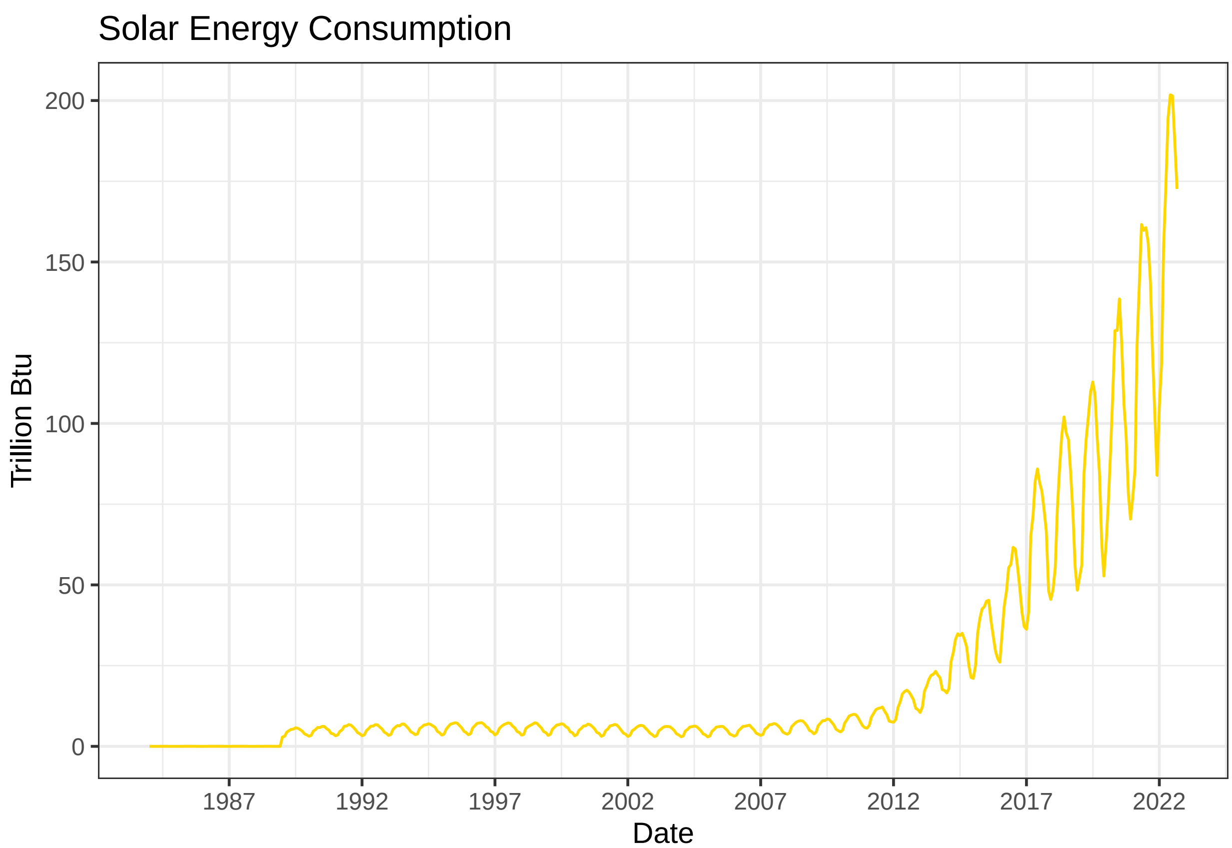
```
# Preserve Orig DF
energy_data_orig <- energy_data

# Filter Data
energy_data <- energy_data_orig %>%
    select(c("Date", "Solar Energy Consumption", "Wind Energy Consumption")) %>%
    rename(Solar.Energy.Consumption = "Solar Energy Consumption",
        Wind.Energy.Consumption = "Wind Energy Consumption") %>%
    filter(Wind.Energy.Consumption != "Not Available") %>%
    filter(Solar.Energy.Consumption != "Not Available") %>%
    mutate(Wind.Energy.Consumption = as.numeric(Wind.Energy.Consumption),
        Solar.Energy.Consumption = as.numeric(Solar.Energy.Consumption))
```

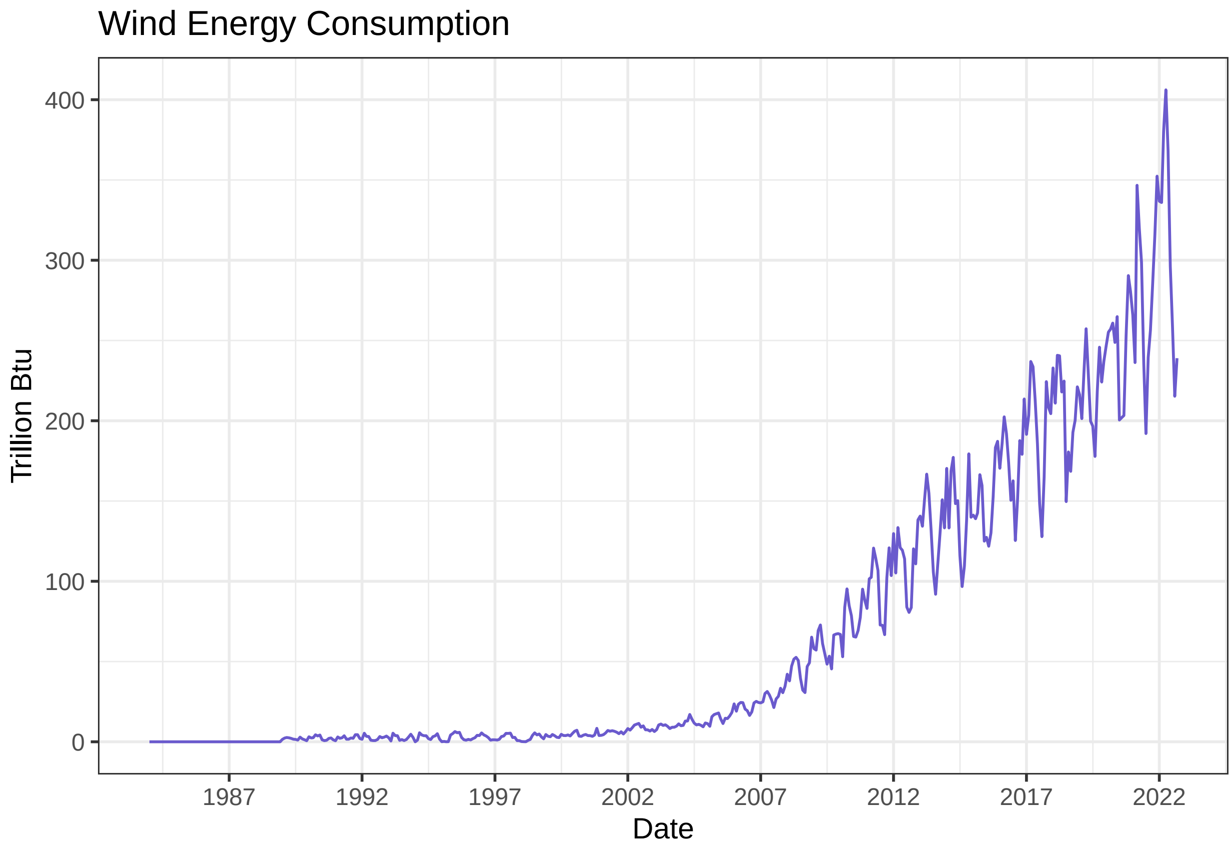
Q2

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")")*

```
# Solar Energy
ggplot(energy_data, aes(x = Date, y = Solar.Energy.Consumption)) +
    geom_line(color = "gold") + ylab("Trillion Btu") + ggtitle("Solar Energy Consumption") +
    scale_x_date(date_breaks = "5 years", date_labels = "%Y") +
    theme_bw()
```



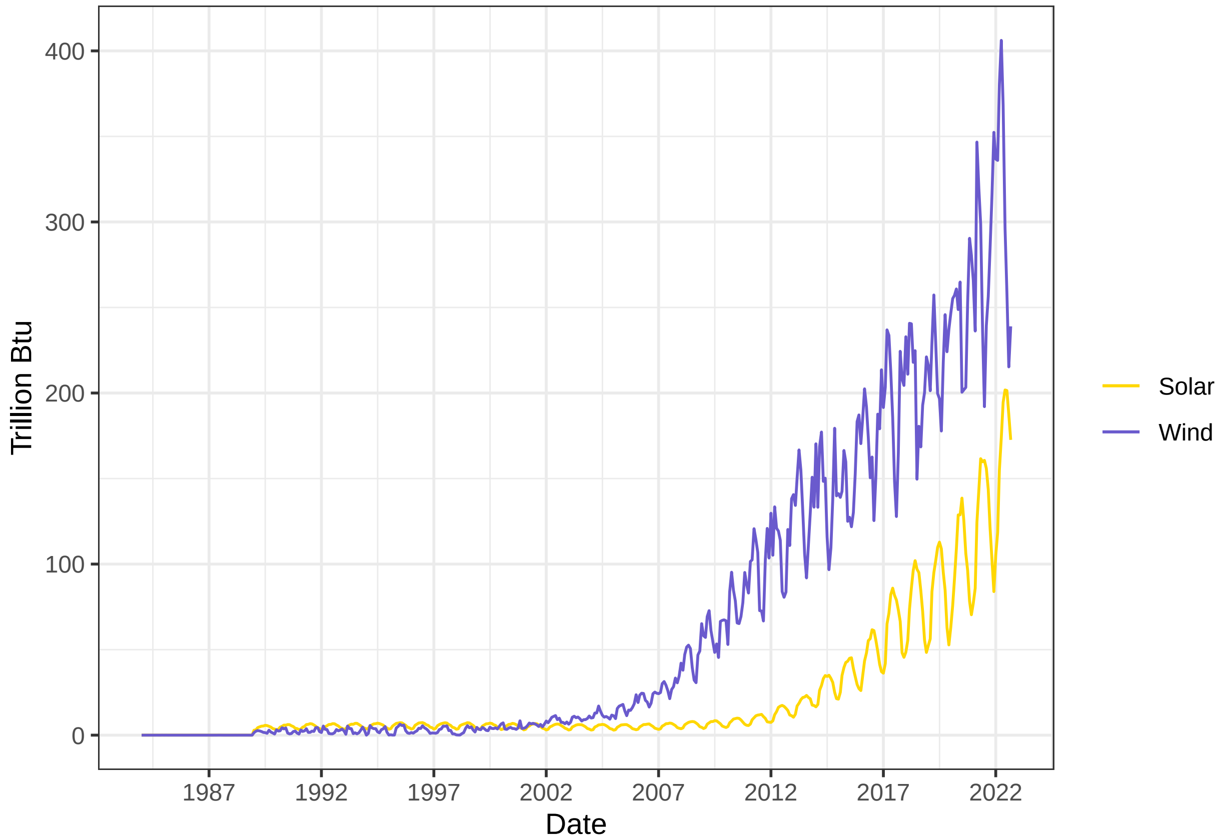

```r
# Wind Energy
ggplot(energy_data, aes(x = Date, y = Wind.Energy.Consumption)) +
    geom_line(color = "slateblue") + ylab("Trillion Btu") + ggtitle("Wind Energy Consumption") +
    scale_x_date(date_breaks = "5 years", date_labels = "%Y") +
    theme_bw()
```



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.

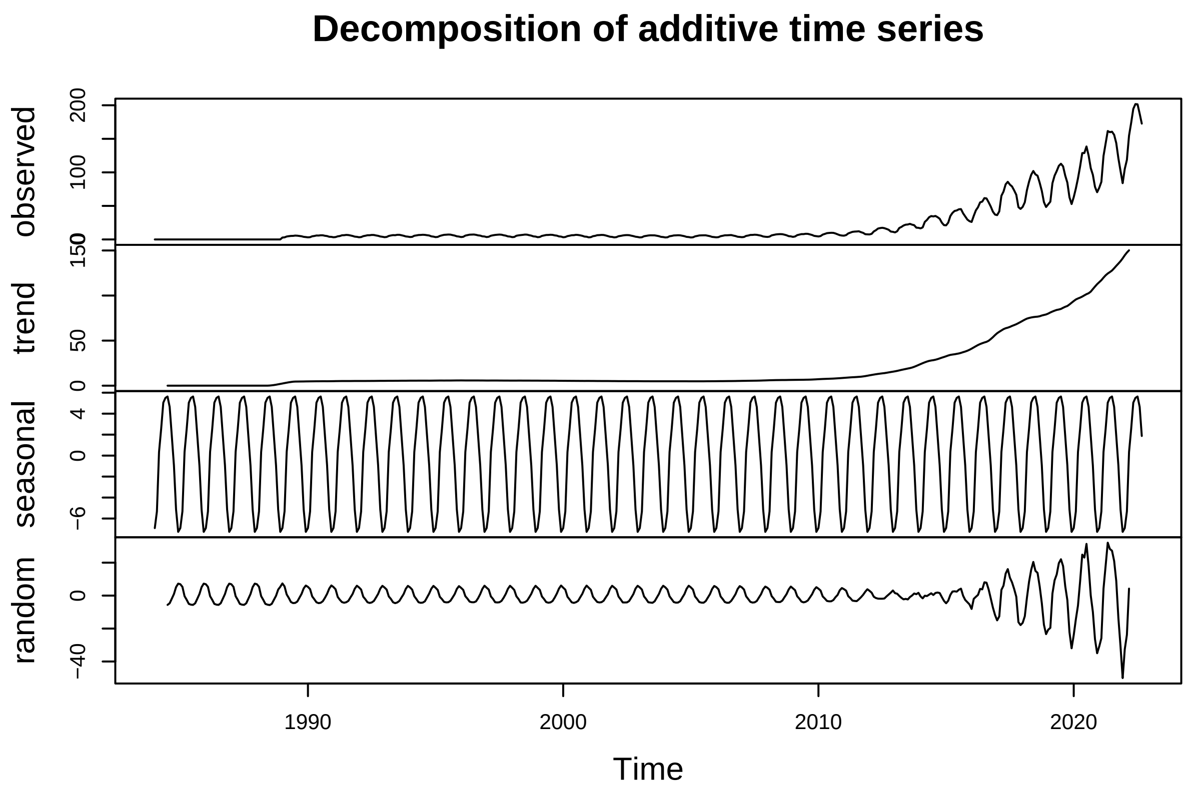
```
# Plot Both Together
ggplot(energy_data) + geom_line(aes(x = Date, y = Solar.Energy.Consumption,
    color = "Solar")) + geom_line(aes(x = Date, y = Wind.Energy.Consumption,
    color = "Wind")) + labs(color = "") + scale_color_manual(values = c(Solar = "gold",
    Wind = "slateblue"), labels = c("Solar", "Wind")) + theme(legend.position = "bottom") +
    ylab(label = "Trillion Btu") + scale_x_date(date_breaks = "5 years",
    date_labels = "%Y") + theme_bw()
```



Q3

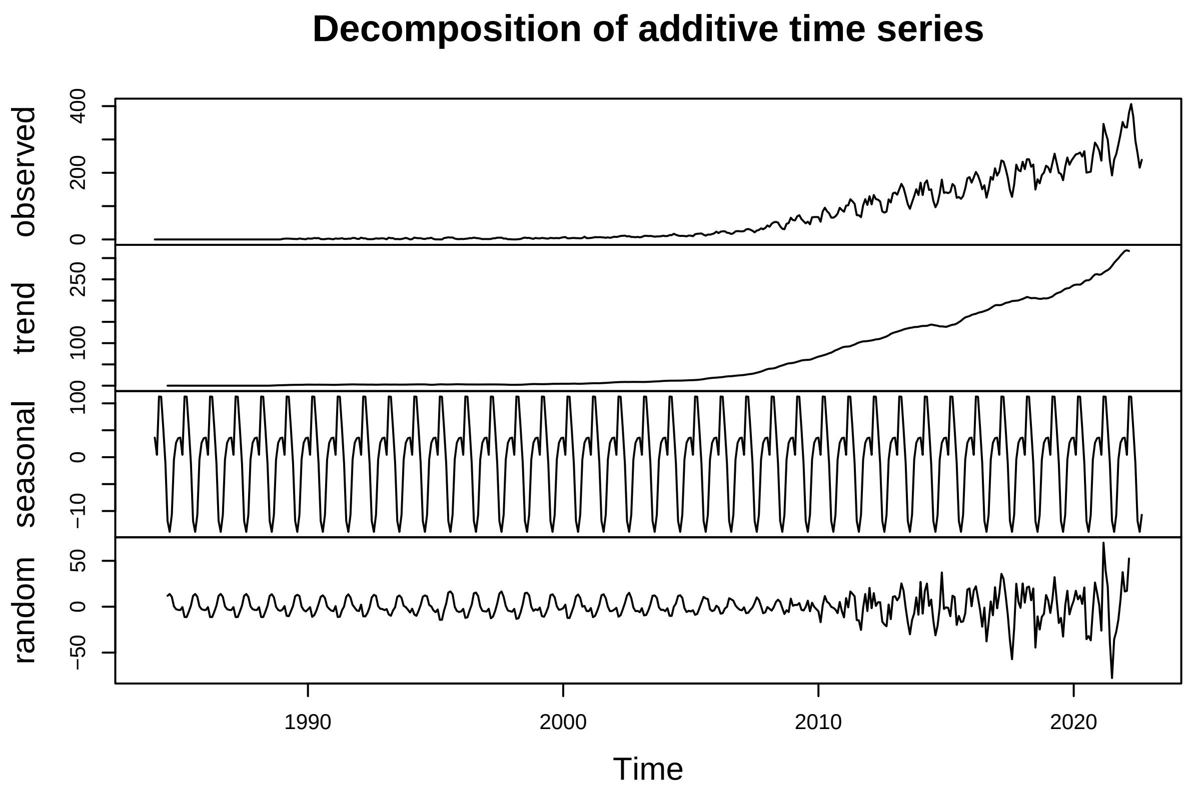
Transform wind and solar series into a time series object and apply the decompose function on them using the additive option, i.e., `decompose(ts_data, type = "additive")`. What can you say about the trend component? What about the random component? Does the random component look random? Or does it appear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

```
# Solar
ts_solar <- ts(energy_data$Solar.Energy.Consumption, start = 1984,
    frequency = 12)
decompose_solar <- decompose(ts_solar, type = "additive")
plot(decompose_solar)
```



The trend component appears to be increasing at an increasing rate over the time period. The random component does not appear to be random, but to have come kind of seasonality or cyclicality still present, demonstrated by the regular wave pattern from the beginning of the series to the mid 2010s, when the wave continues but the magnitude increases.

```
# Wind
ts_wind <- ts(energy_data$Wind.Energy.Consumption, start = 1984,
    frequency = 12)
decompose_wind <- decompose(ts_wind, type = "additive")
plot(decompose_wind)
```

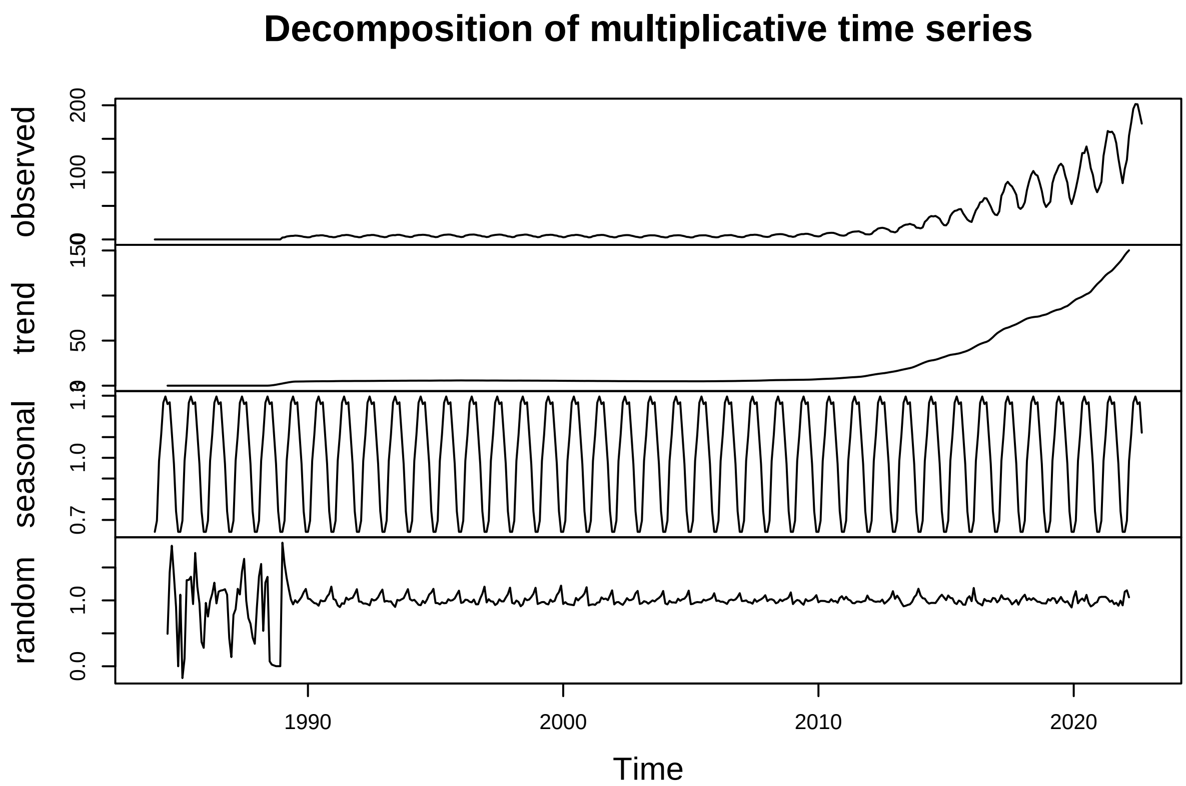


The decomposition for wind is similar to the decomposition for solar. The trend starts increasing at a relatively constant rate in the late 2000s. The random component still demonstrates some seasonality, although this follows a slightly different pattern with 3-4 seasonal components (a peak followed by a small decrease and then a distinct valley) compared to the smooth up and down smooth of the solar trend and random component. I expect that the solar seasonality reflects the longer and shorter days based on the annual rotation of the earth around the sun, while the wind follows some kind of distinct pattern across all four seasons.

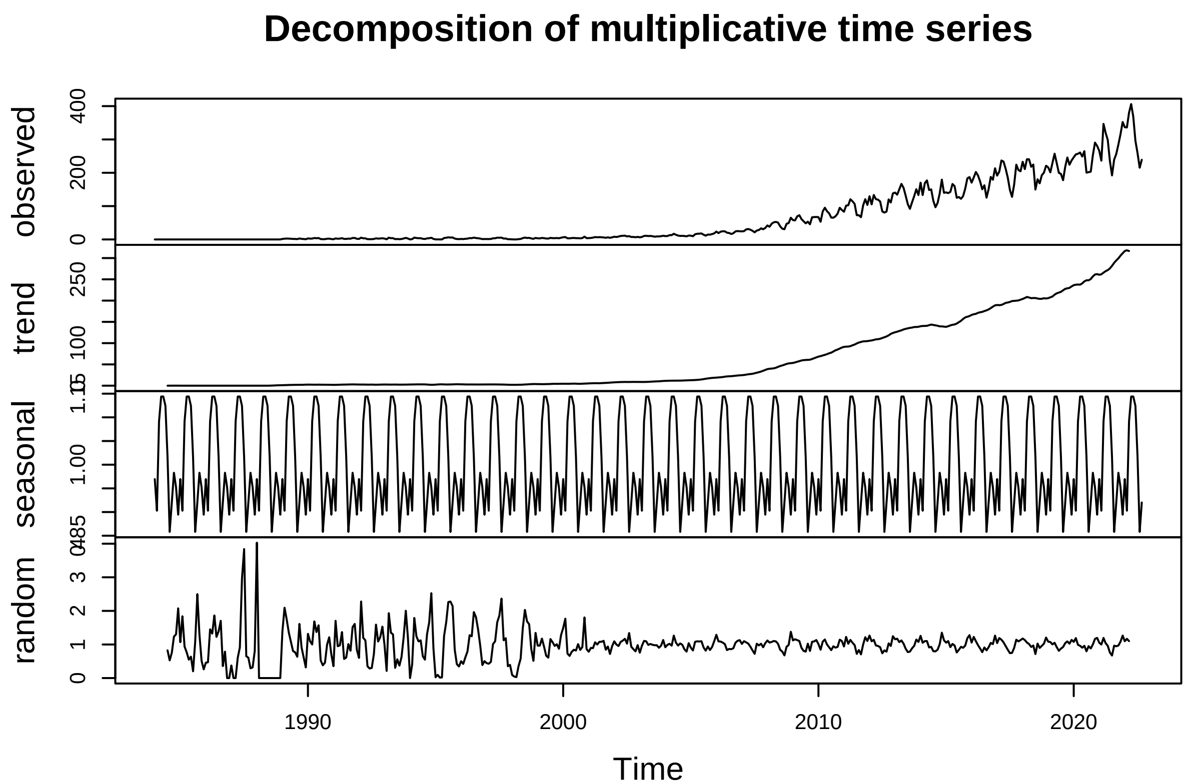
Q4

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?

```
# Solar Multiplicative
decompose_solar_mult <- decompose(ts_solar, type = "multiplicative")
plot(decompose_solar_mult)
```




```r
# Wind Multiplicative
decompose_wind_mult <- decompose(ts_wind, type = "multiplicative")
plot(decompose_wind_mult)
```



In the multiplicative decomposition, the random component looks much more random. However, the random component looks best during the periods of time before each series started increasing. After the series start increasing, the random component becomes much smaller in magnitude and still exhibits some regular peaks and valleys that indicate some sort of cyclicality that isn't accounted for completely.

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.

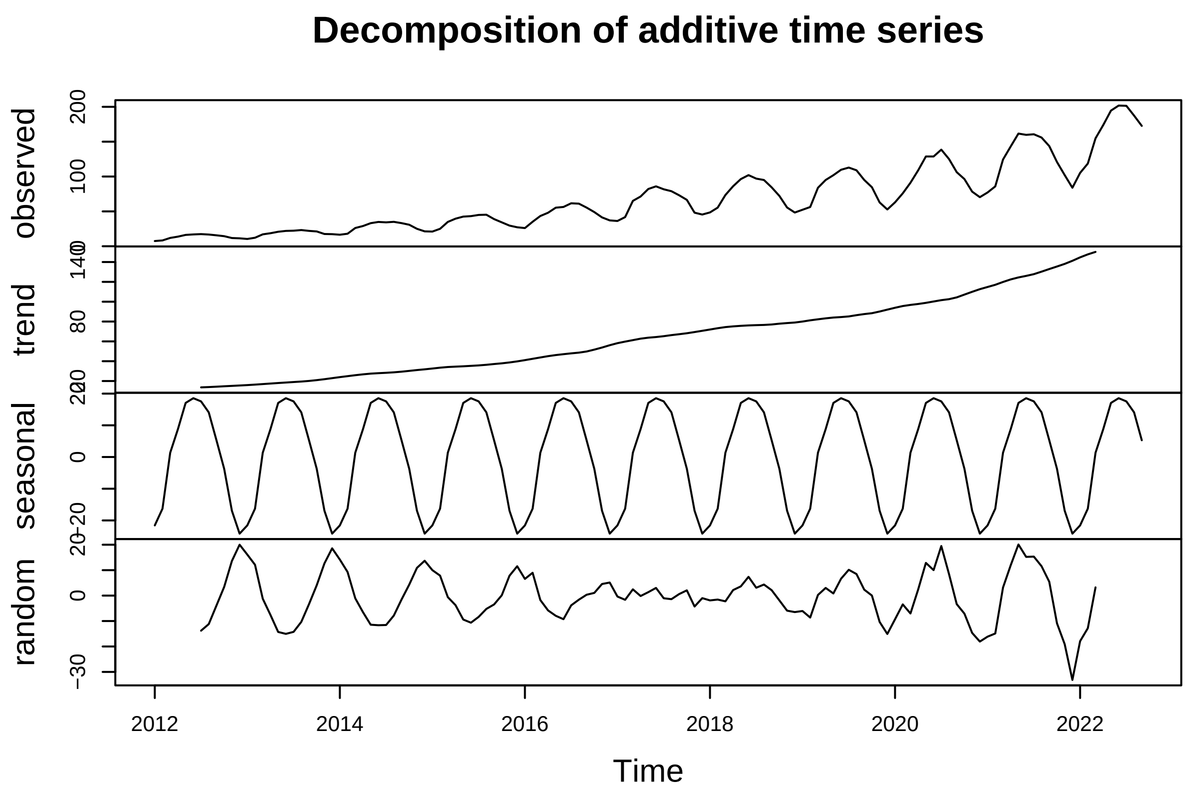
> Answer: The amount of data needed depends on the purpose and time scale of the forecast. Although solar and wind technologies have been around for a long time, they have only been implemented on a large scale in recent years, while we can see clearly in their trends. For wind, consumption started increasing regularly in the late 2000s, while for solar it start increasing in the early 2010s. If we are just interested in explaining the next six months of solar and wind consumption, then we do not need all the data from before these time frames. I think it would be sufficient to consider only what has happened since the significant and seemingly permanent change in trend occured for each series. However, if we were interested in forecasting over a much longer time frame, like the next 40 years, then we might need to take more of the data into consideration and attempt to understand how the trends responded to specific technology and policy changes.

Q6

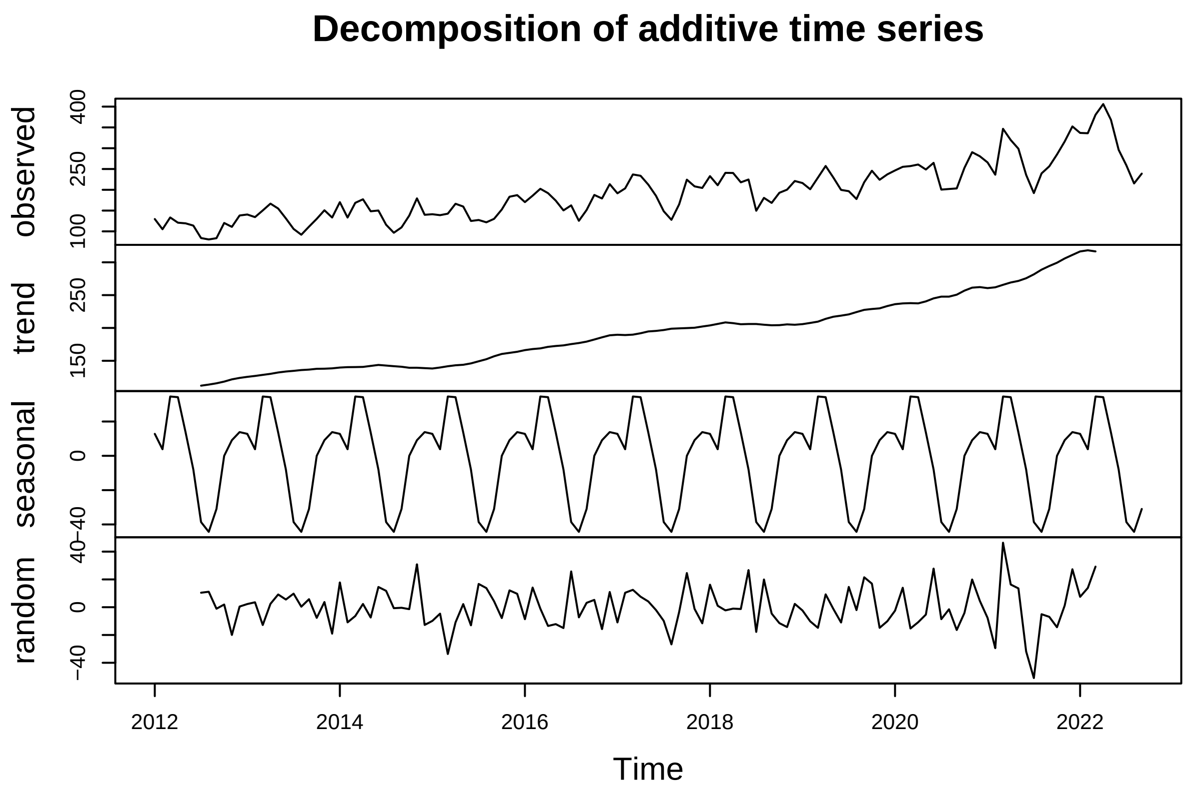
Create a new time series object where historical data starts on January 2012. Hint: use `filter()` function so that you don't need to point to row numbers, .i.e, `filter(xxxx, year(Date) >= 2012 )`. Apply the decompose function `type=additive` to this new time series. Comment the results. Does the random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

```
energy_data_recent <- energy_data %>%
    filter(year(Date) >= 2012)

# Solar
ts_solar_recent <- ts(energy_data_recent$Solar.Energy.Consumption,
    start = 2012, frequency = 12)
decompose_solar_recent <- decompose(ts_solar_recent, type = "additive")
plot(decompose_solar_recent)
```




```
# Wind
ts_wind_recent <- ts(energy_data_recent$Wind.Energy.Consumption,
    start = 2012, frequency = 12)
decompose_wind_recent <- decompose(ts_wind_recent, type = "additive")
plot(decompose_wind_recent)
```



Answer: For the solar consumption series, the random component still demonstrats a cyclicality. In earlier years, this mirrors the seasonal component (the random component spikes when the seasonal component is at its lowest point), looks more random during the middle of the series, and follows the seasonal components (spikes when the seasonal component spikes) at the end of the series. This likely reflects the change in magnitude of the seasonality over the series, as seen in the "observed" panel of the decomposition plot. As solar consumption increases, annual variation in solar radiation affects all solar energy sources and so the seasonal effect becomes more prominent as the trend increases, so maybe the multiplicative model would be better for this series. For the wind series, limiting to 2012 and later and using the additive decomposition seems to really eliminate any remaining seasonality or cyclicality in the random component.