

DATA 607 - Homework 1

OMER OZEREN

Table of Contents

Question 2.1.....	1
Gold	2
Wollyrnq	2
Gasoline	3
Question 2.2.....	4
A	4
B.....	4
C.....	5
Question 2.3.....	6
Question 2.6.....	10
hsales.....	10
usdeaths	13
bricksq	16
sunspotarea	19
gasoline.....	21

Exercises 2.1, 2.2, 2.3 and 2.6 from the Hyndman online Forecasting book.

#Load req's packages

```
library(forecast)
```

```
library(readxl)
```

```
library(RCurl)
```

```
library(fpp2)
```

Question 2.1

Use the help function to explore what the series *gold*, *woollyrnq* and *gas* represent.

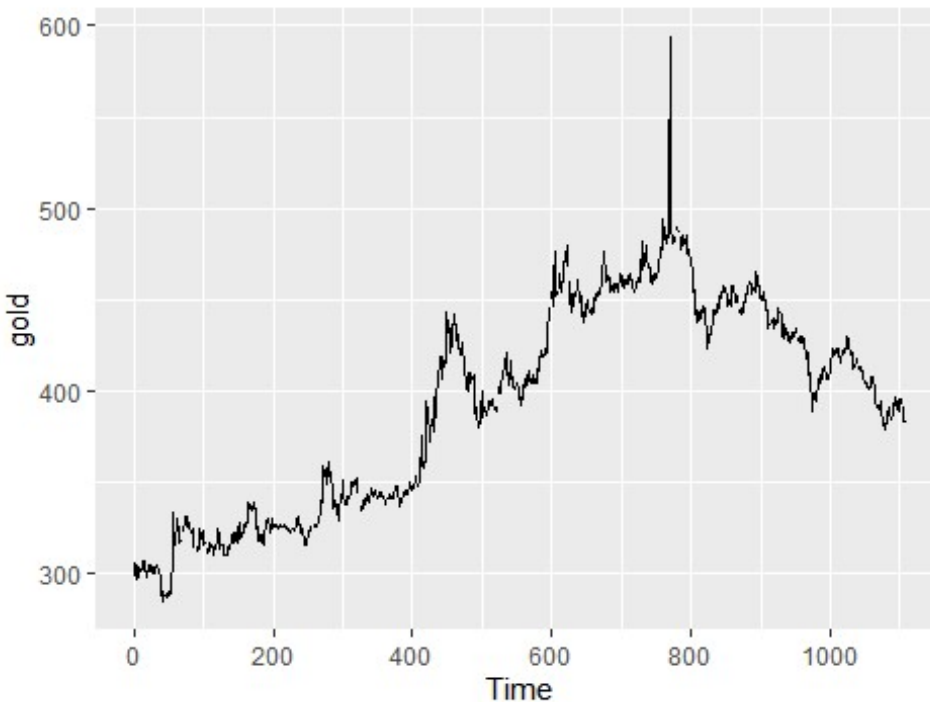
- Use `autoplot()` to plot each of these in separate plots.
- What is the frequency of each series? Hint: apply the `frequency()` function.
- Use `which.max()` to spot the outlier in the gold series. Which observation was it?

```
describe.data <- function(data) {  
  freq <- frequency(data)
```

```
outlier <- which.max(data)
return(c(freq,outlier))
}
```

Gold

```
autoplot(gold)
```

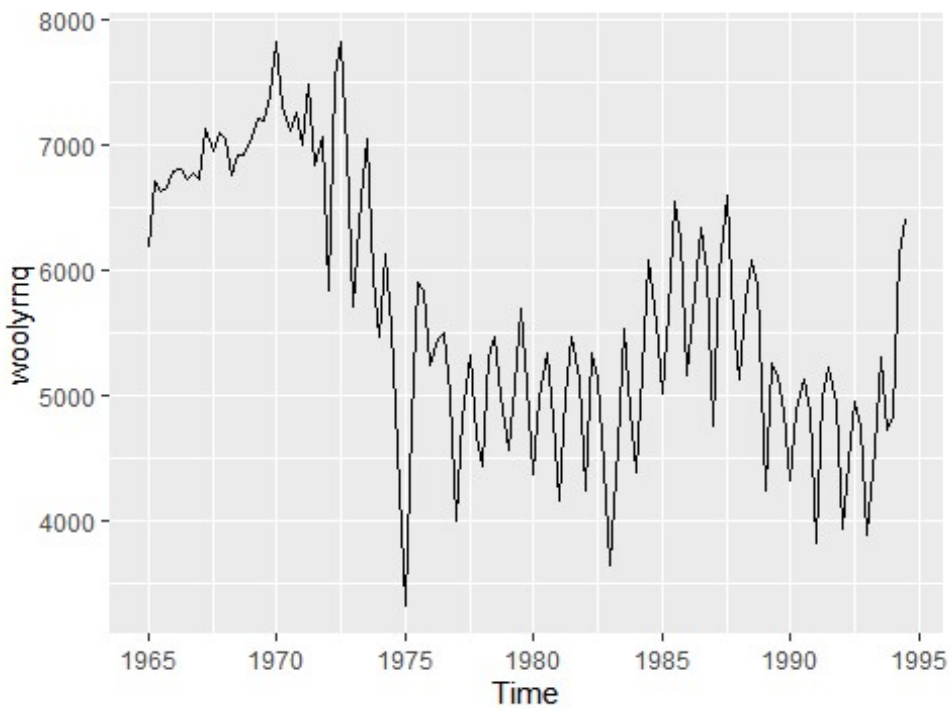


```
question1 <- describe.data(gold)
```

- The “gold” data represents the daily (morning) gold price in USD for the period 1985-01-01 to 1989-03-31.
- The frequency of this data is 1. The data is daily.
- The outlier datapoint appears at index 770 and corresponds to a price of 593.7

Wollyrnq

```
autoplot(woollyrnq)
```

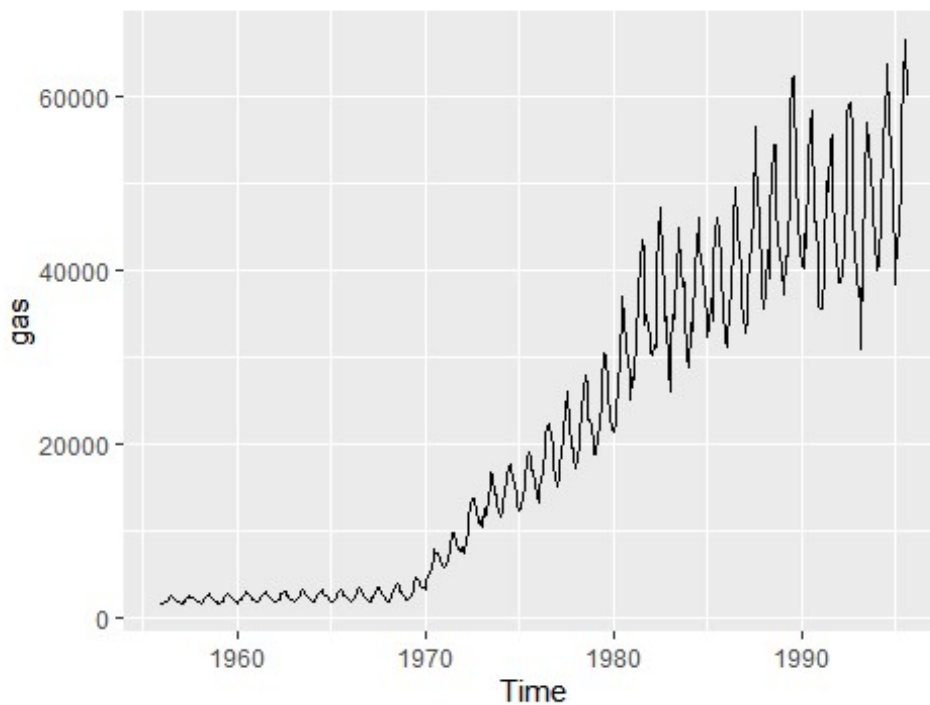


```
question2 <- describe.data(woolyrnq)
```

- The “woolyrnq” data represents the quarterly production of woolen yard (tonnes) in Australia for the period Mar-1965 to Sep-1994
- The frequency of this data is 4 observation per year, i.e. quarterly.

Gasoline

```
autoplot(gas)
```



```
question3 <- describe.data(gas)
```

- The “gas” dataset shows Australian monthly gas production (units not specified) for the period 1956-1995
- The frequency of this data is 4 observation per year, i.e. monthly.

Question 2.2

Download the file *tute1.csv* from the book website, open it in Excel (or some other spreadsheet application), and review its contents. You should find four columns of information. Columns B through D each contain a quarterly series, labelled Sales, AdBudget and GDP. Sales contains the quarterly sales for a small company over the period 1981-2005. AdBudget is the advertising budget and GDP is the gross domestic product. All series have been adjusted for inflation.

A

You can read the data into R with the following script:

```
tute1 <- read.csv("http://otexts.com/fpp2/extrafiles/tute1.csv", header=T)
View(tute1)
```

B

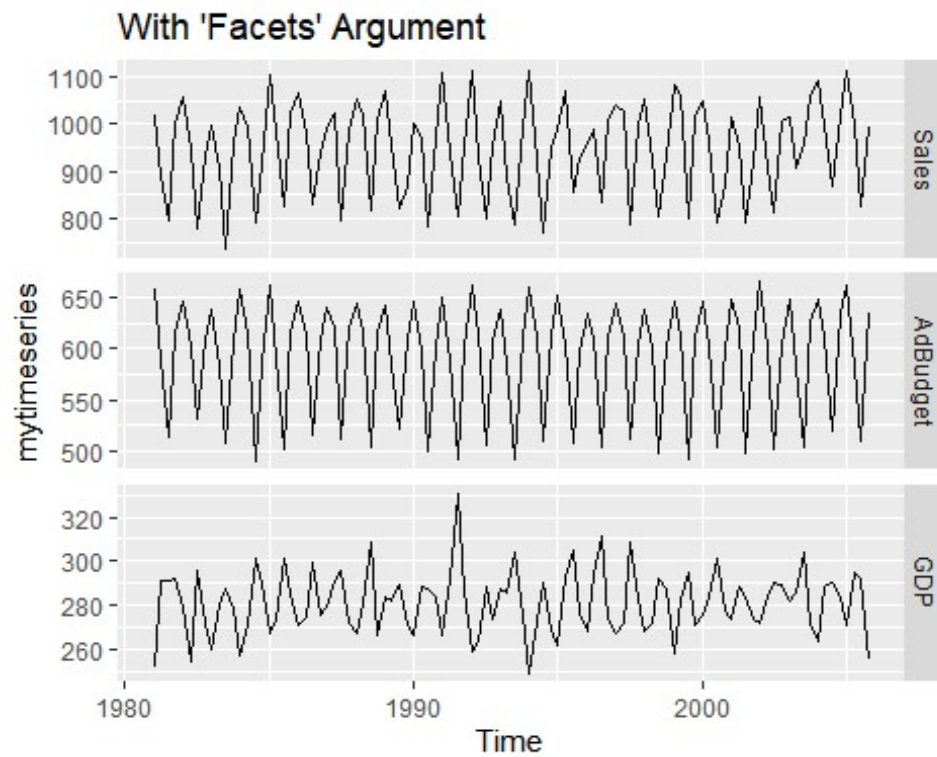
Convert the data to time series

```
mytimeseries <- ts(tute1[, -1], start=1981, frequency=4)
```

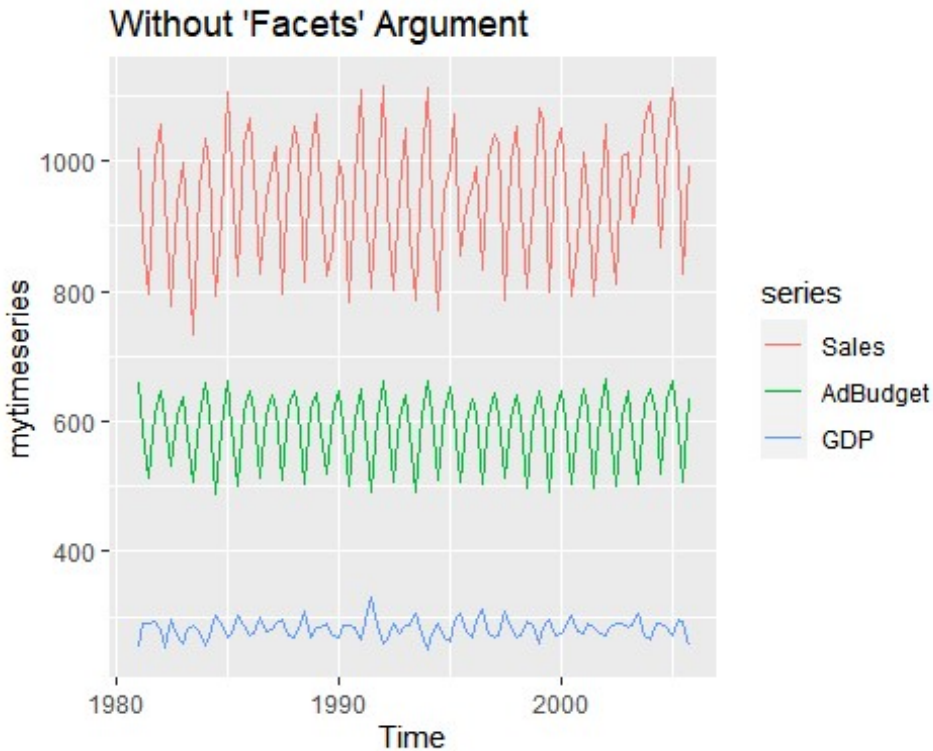
C

Construct time series plots of each of the three series & check what happens when you don't include `facets=TRUE`.

```
autoplot(mytimeseries, facets=TRUE, main="With 'Facets' Argument")
```



```
autoplot(mytimeseries, main="Without 'Facets' Argument")
```

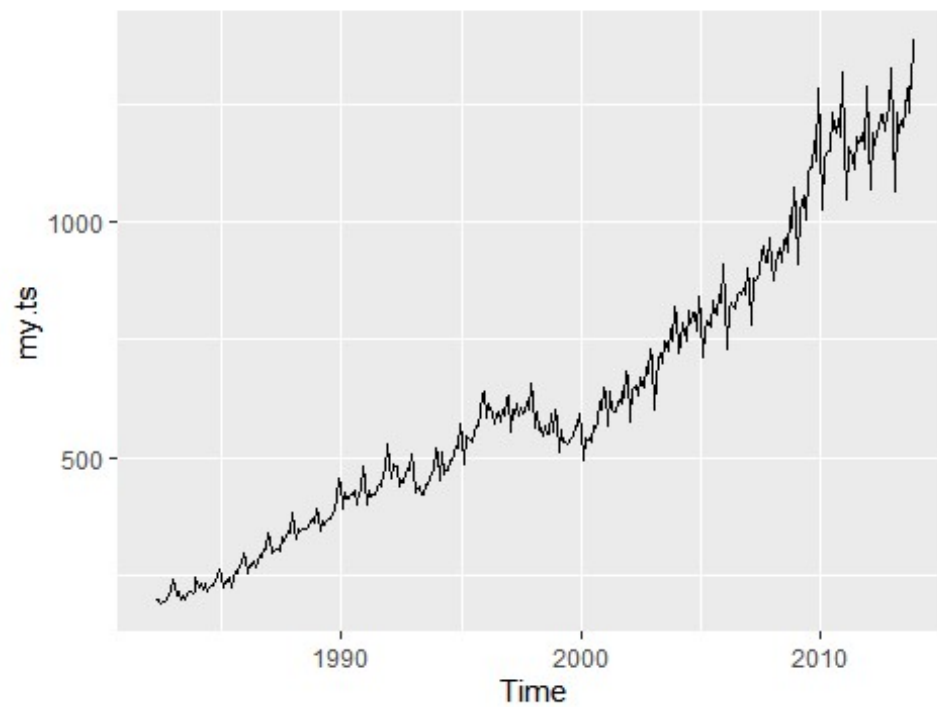


Question 2.3

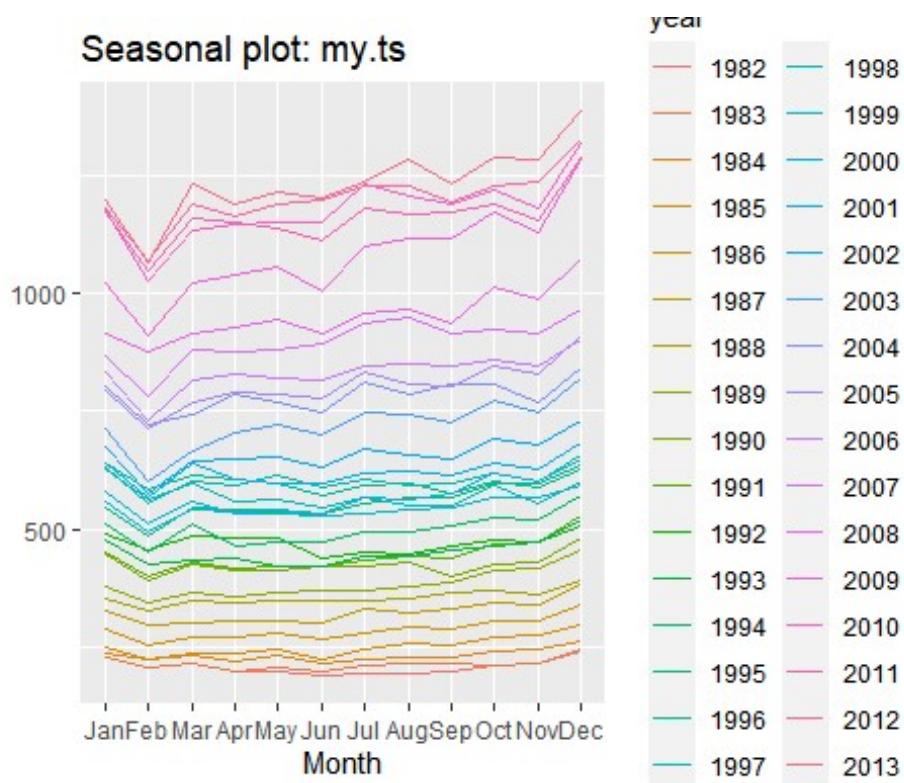
Download some monthly Australian retail data from the book website. These represent retail sales in various categories for different Australian states, and are stored in a MS-Excel file.

- You can read the data into R with the following script:
- Select one of the time series as follows (but replace the column name with your own chosen column):
- Explore your chosen retail time series using the following functions:

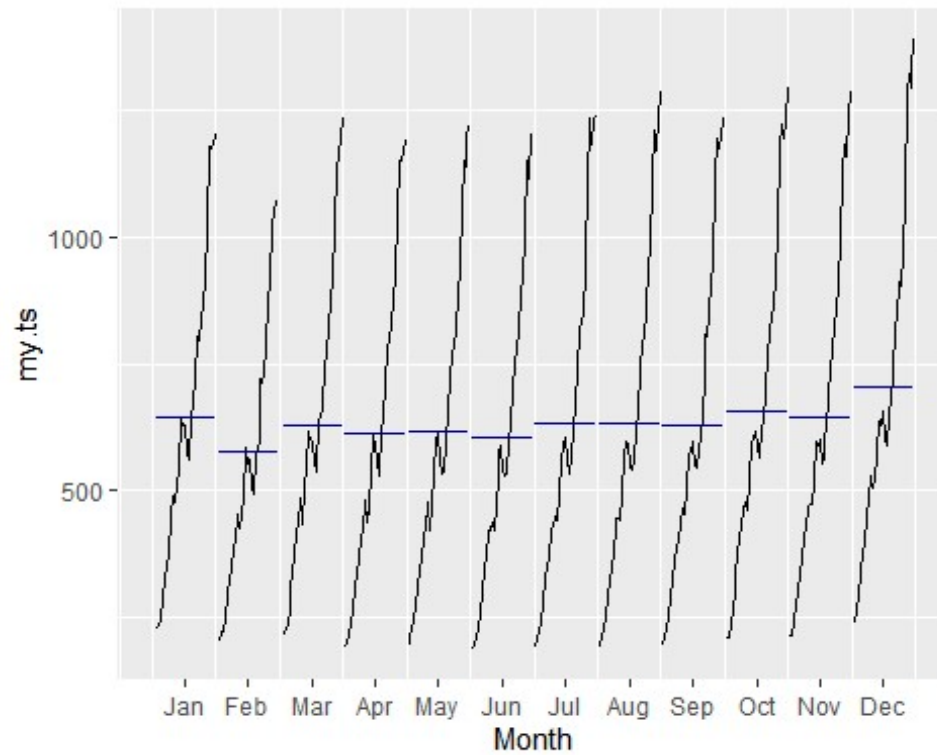
```
#create a temp file
temp_file <- tempfile(fileext = ".xlsx")
download.file(url =
  "https://github.com/omerozeren/DATA624/raw/master/HMW1/retail.xlsx",
              destfile = temp_file,
              mode = "wb",
              quiet = TRUE)
#Load xl from temp
retaildata <- readxl::read_excel(temp_file, skip=1)
my.ts <- ts(retaildata[, "A3349388W"],
            frequency=12, start=c(1982,4))
autoplot(my.ts)
```



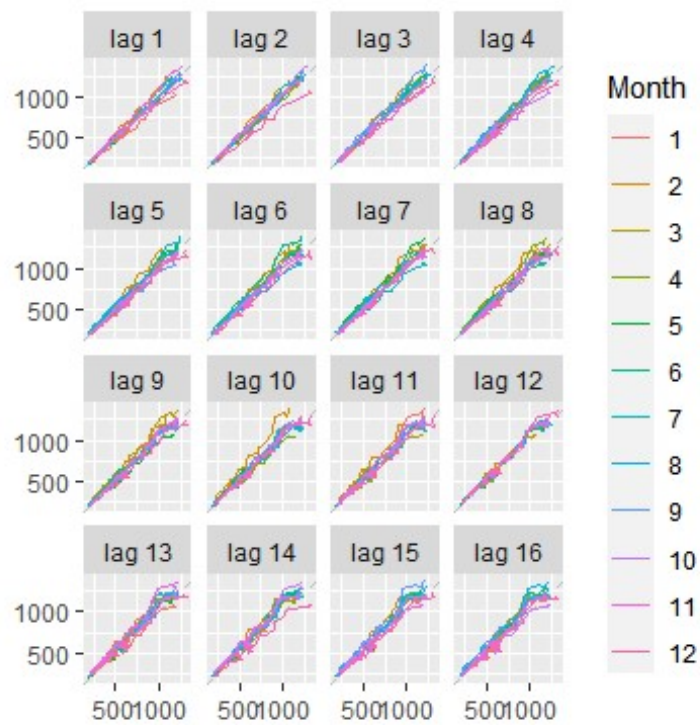
```
ggseasonplot(my.ts)
```



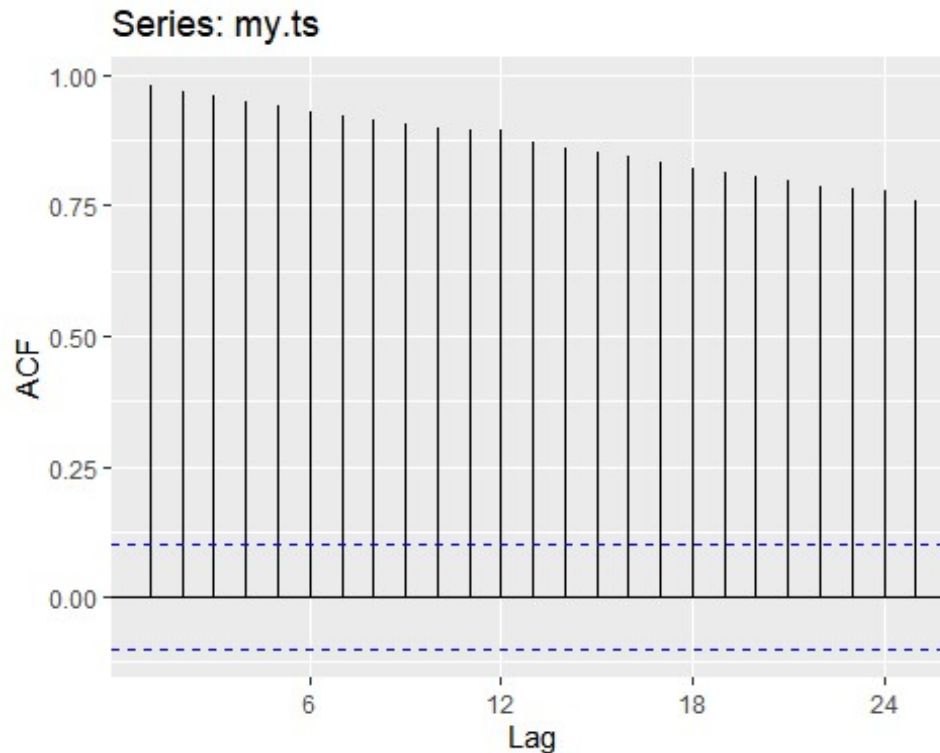
```
ggsubseriesplot(my.ts)
```



`gglagplot(my.ts)`



`ggAcf(my.ts)`



Reference:

<https://otexts.com/fpp2/autocorrelation.html>

“Trend: A trend exists when there is a long-term increase or decrease in the data. It does not have to be linear. Sometimes we will refer to a trend as “changing direction”, when it might go from an increasing trend to a decreasing trend.”

“Seasonal: A seasonal pattern occurs when a time series is affected by seasonal factors such as the time of the year or the day of the week. Seasonality is always of a fixed and known frequency. The monthly sales of antidiabetic drugs above shows seasonality which is induced partly by the change in the cost of the drugs at the end of the calendar year.”

“Cyclic: A cycle occurs when the data exhibit rises and falls that are not of a fixed frequency. These fluctuations are usually due to economic conditions, and are often related to the “business cycle”. The duration of these fluctuations is usually at least 2 years.”

“When data have a trend, the autocorrelations for small lags tend to be large and positive because observations nearby in time are also nearby in size. So the ACF of trended time series tend to have positive values that slowly decrease as the lags increase.

When data are seasonal, the autocorrelations will be larger for the seasonal lags (at multiples of the seasonal frequency) than for other lags.

When data are both trended and seasonal, you see a combination of these effects.”

For this question I just picked a column at random and ended up with “Turnover ; Total (State) ;

- As per the timeseries plot, there is a clear upward sloping trend in the data and there are definitely some seasonal / cyclical effects apparent
- The seasonality plot shows a few things: a consistent dip in Feb, a rise towards the end of the year.
- The ACF plot shows a high degree of auto-correlation with low decay.

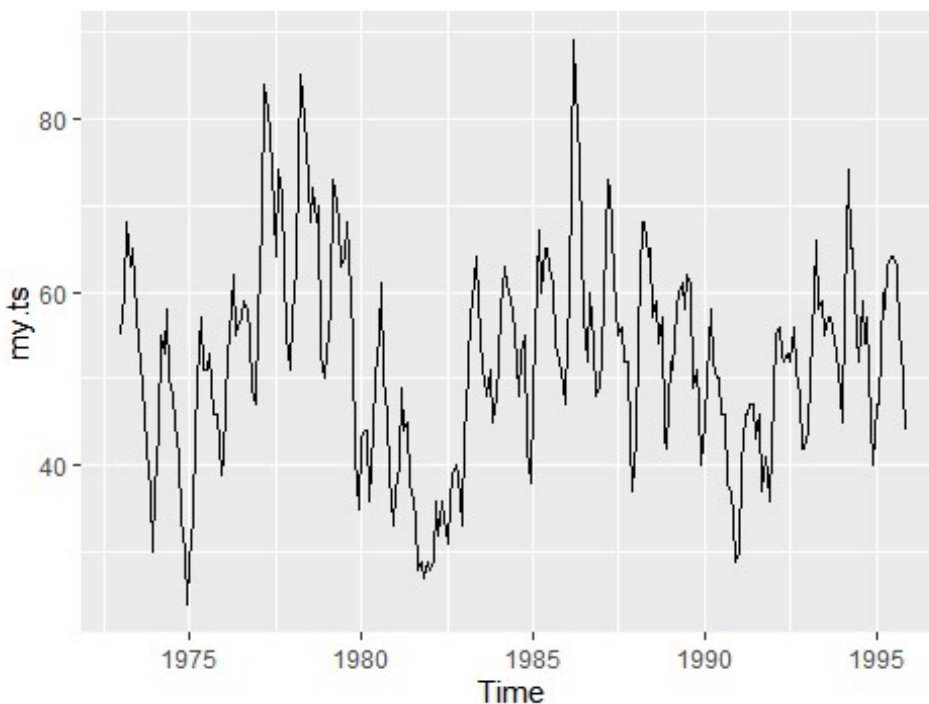
Question 2.6

Use the following graphics functions: `autoplot()`, `ggseasonplot()`, `ggsubseriesplot()`, `gglagplot()`, `ggAcf()` and explore features from the following time series: `hsales`, `usdeaths`, `bricksq`, `sunspotarea`, `gasoline`.

Can you spot any seasonality, cyclicity and trend? What do you learn about the series?

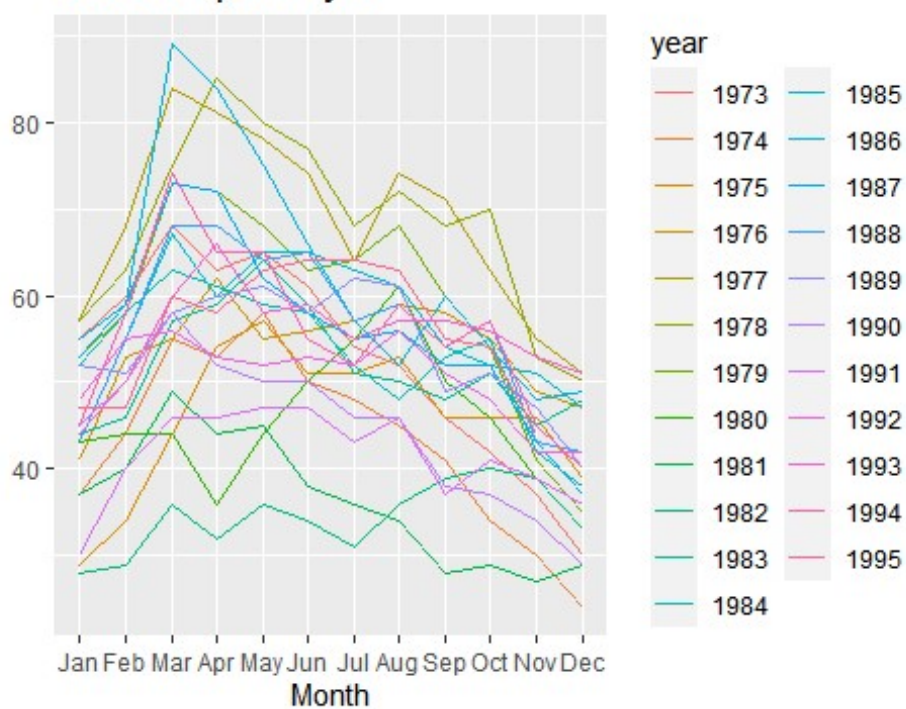
hsales

```
my.ts <- hsales
autoplot(my.ts)
```

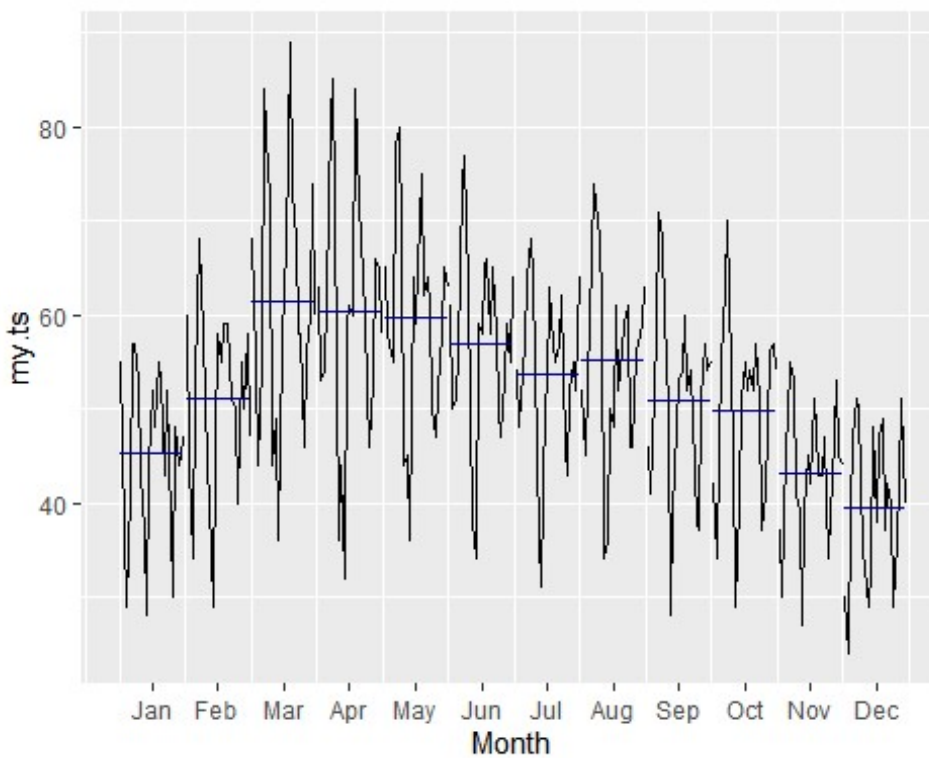


```
ggseasonplot(my.ts)
```

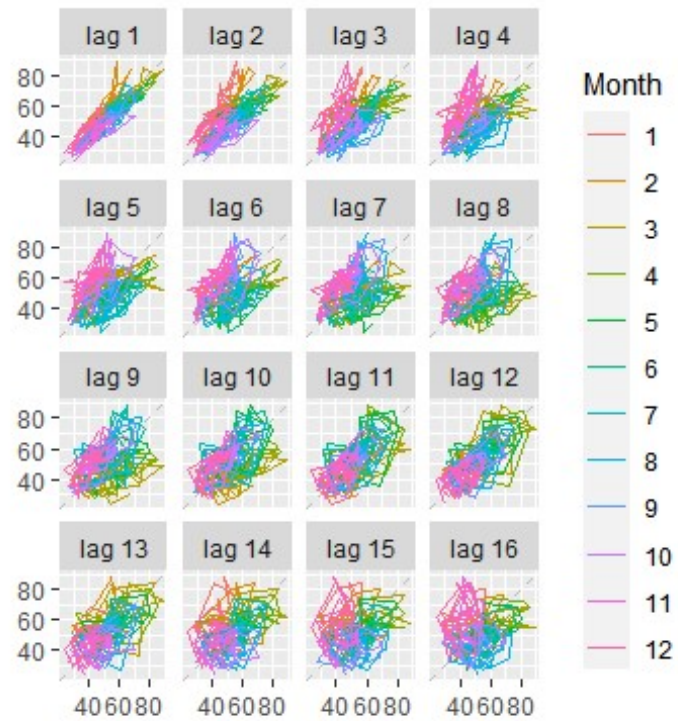
Seasonal plot: my.ts



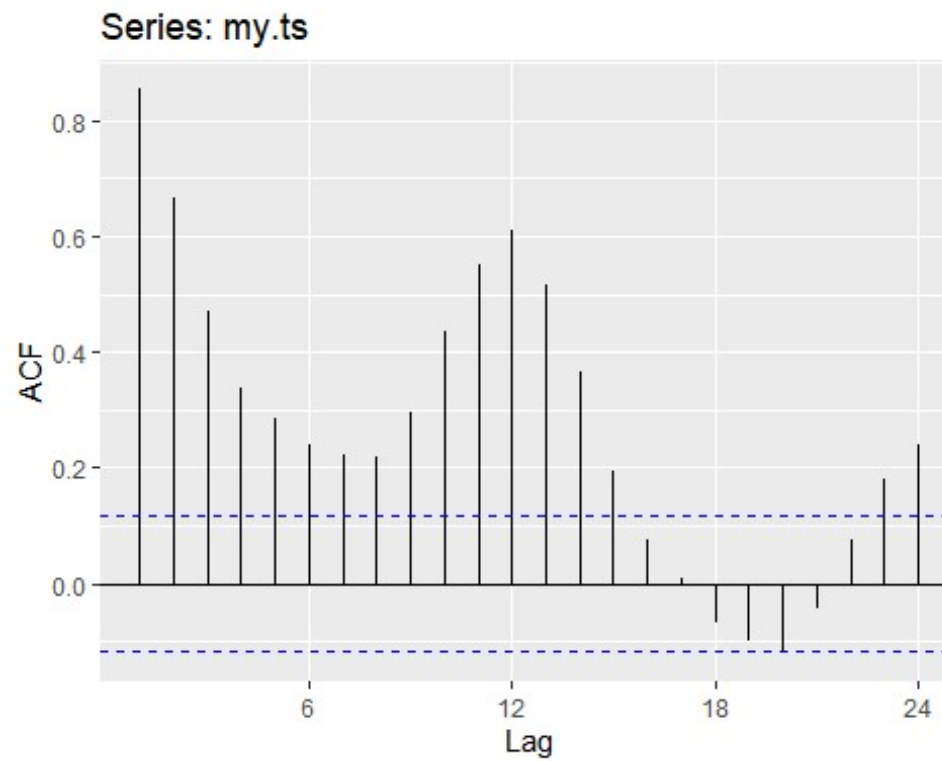
`ggsubseriesplot(my.ts)`



`gglagplot(my.ts)`



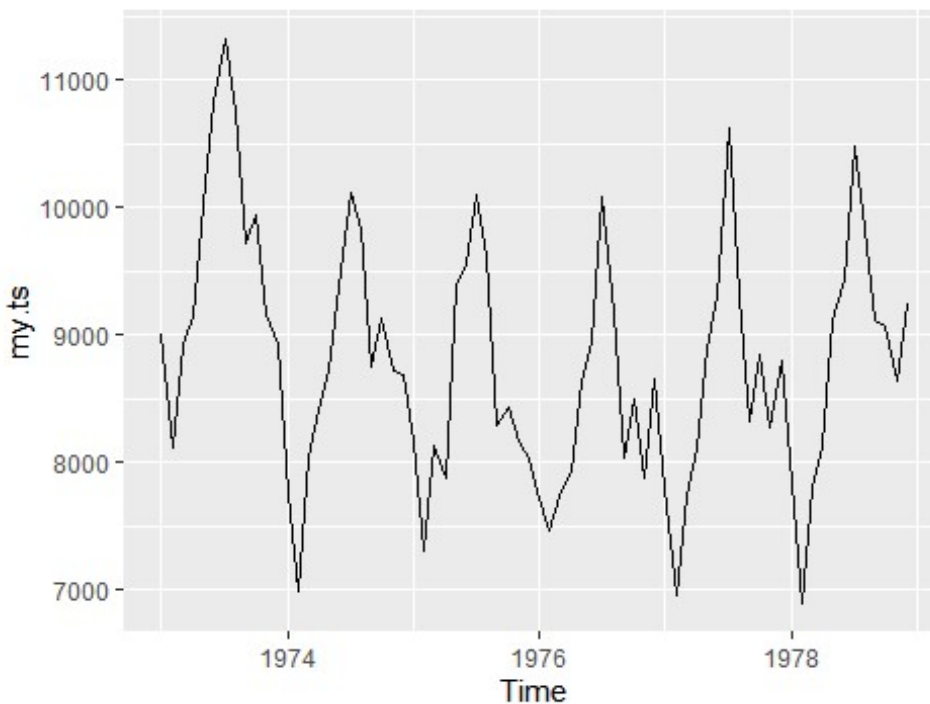
`ggAcf(my.ts)`



- As per the timeseries plot, there is no obvious trend, but it appears as though there may be some seasonality, and there absolutely appears to be some cyclicity - large long-term oscillations between 30 and 80.
- The seasonality plots show a an upward effect between Jan and March followed by a downward slope for the remainder of the year. This is more apparent in the seasonality plot - the subseries plots don't work too well for this data.
- The ACF plot show a high degree of autocorrelation for short lookbacks (1-2 periods) and appear to capture some of the seasonal pattern thereafter.

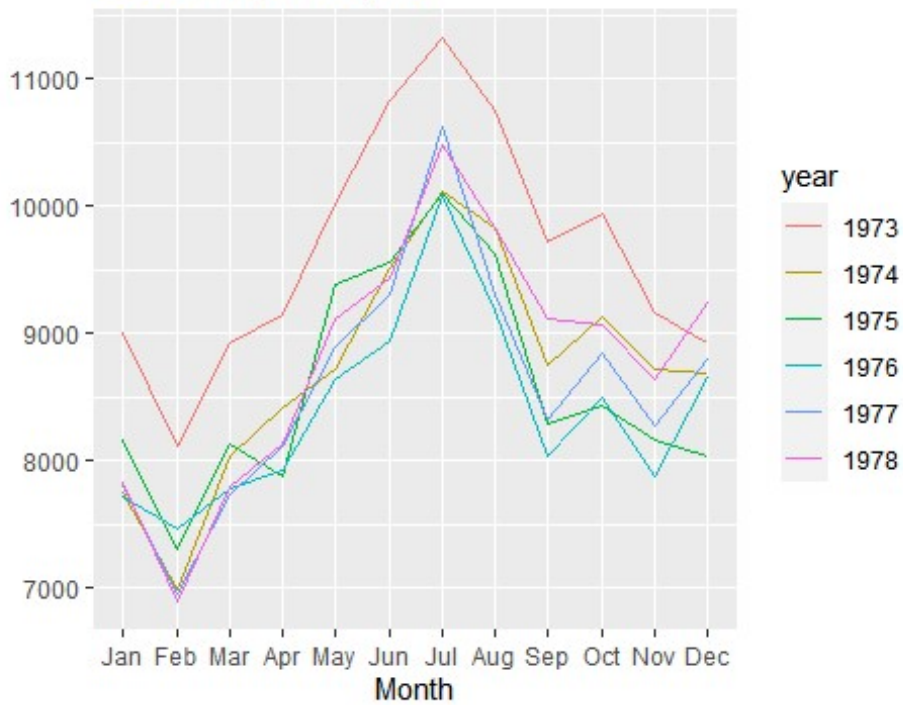
usdeaths

```
my.ts <- usdeaths  
autoplot(my.ts)
```

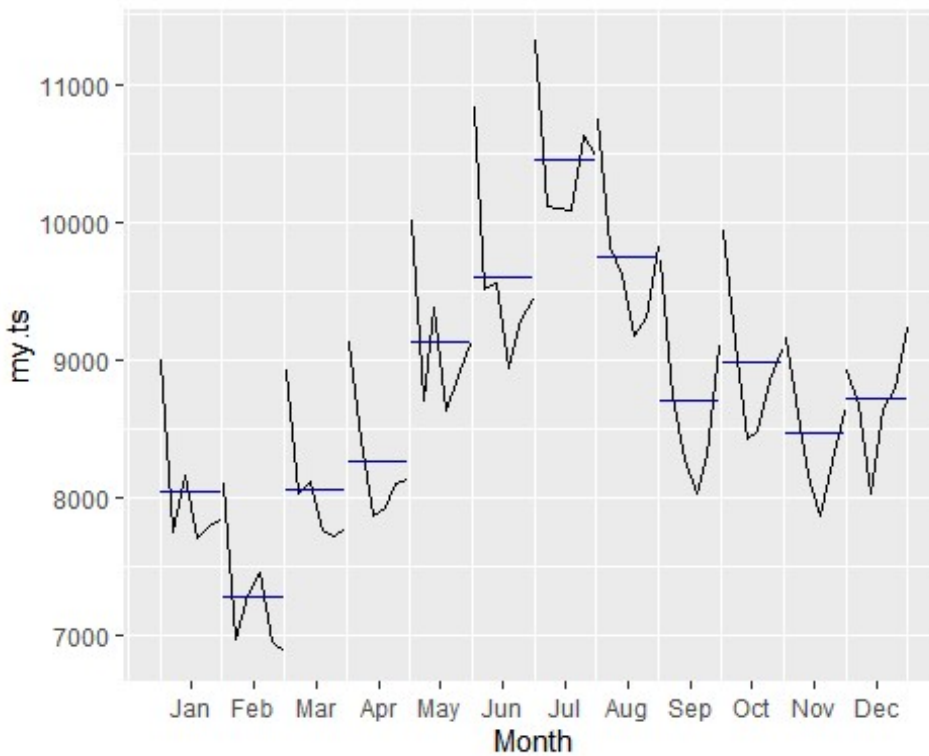


```
ggseasonplot(my.ts)
```

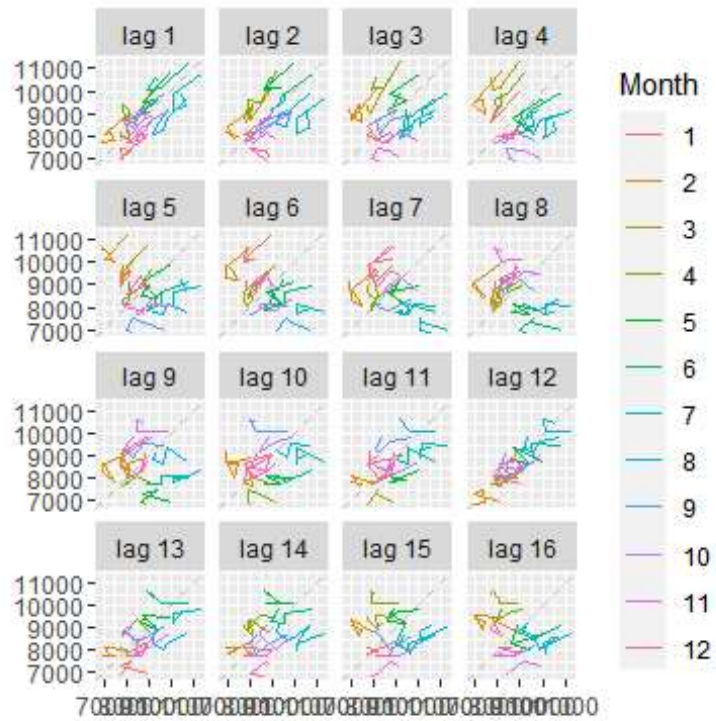
Seasonal plot: my.ts



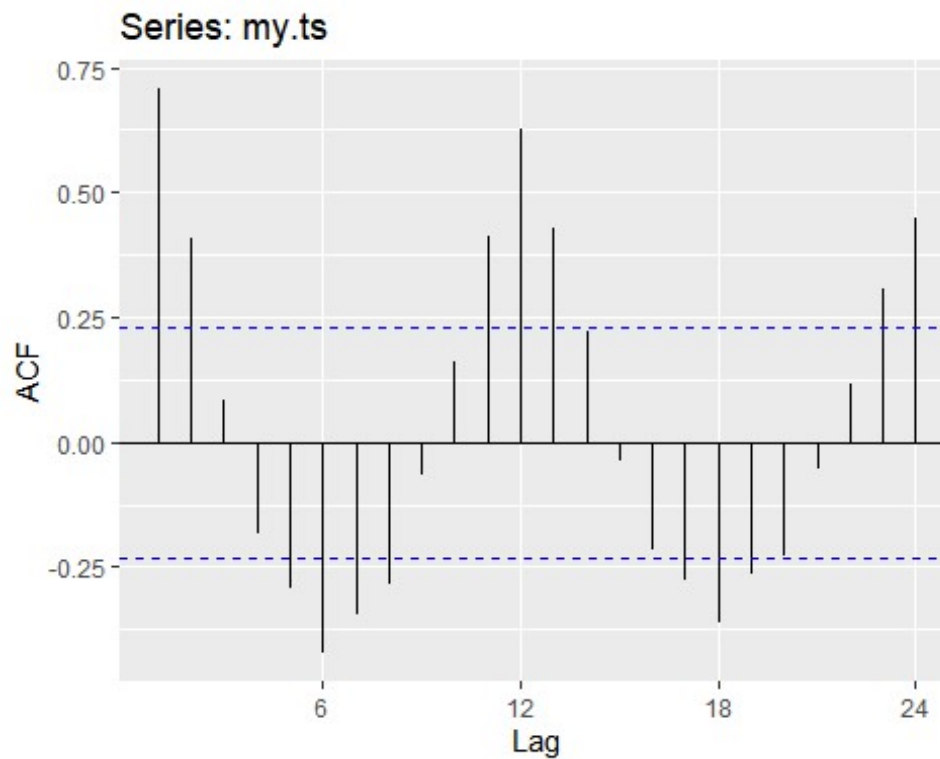
`ggsubseriesplot(my.ts)`



`gglagplot(my.ts)`



`ggAcf(my.ts)`

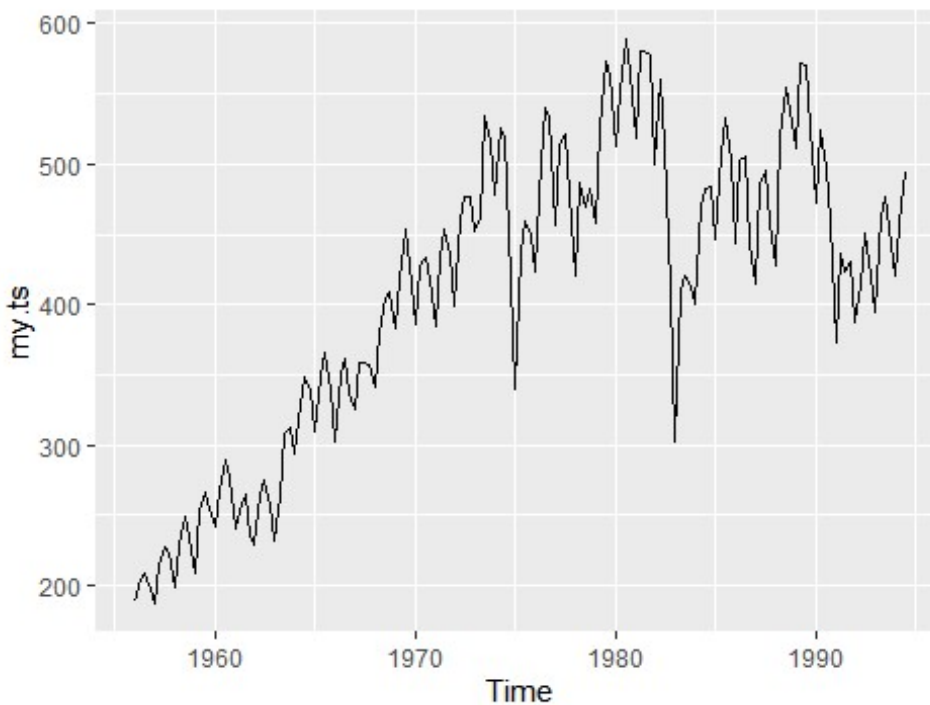


- The data is clearly seasonal and shows little trend, on aggregate.

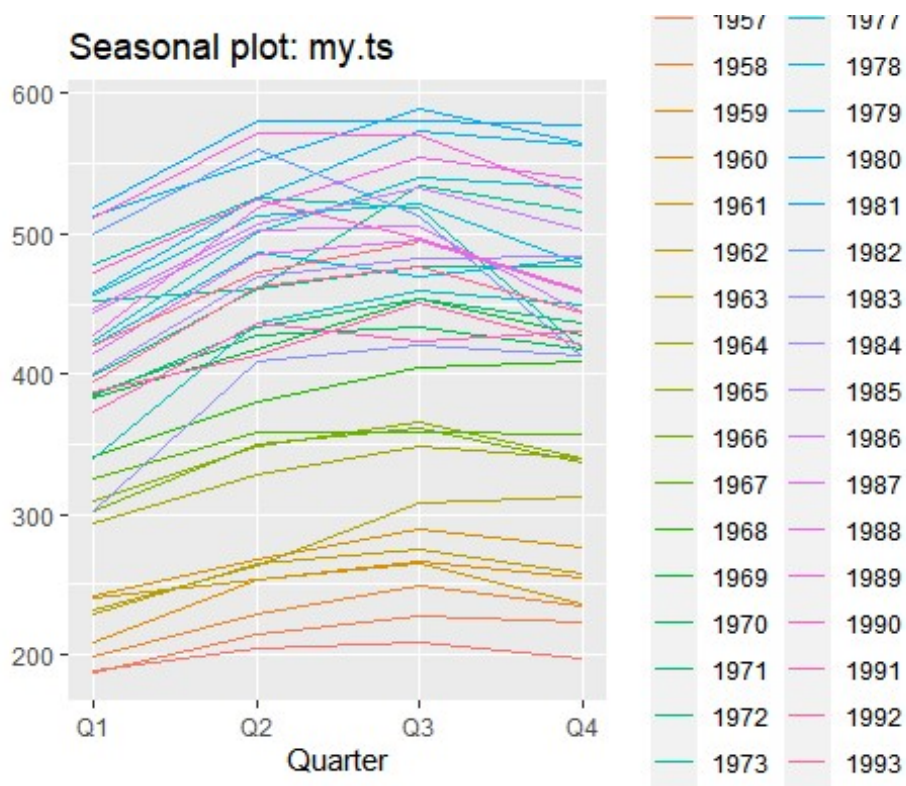
- The seasonality plots shows a dip in Feb, then a clear rise to a peak in Jul. Thereafter, it drops off slightly and seems to flatten out towards the end of the year.
- The lag plots are informative here also with Feb consistently appearing at the bottom and Jul, near the top. The 12 month panel suggests an annual seasonality.
- The ACF is indicative of a pattern in the data. The peaks are at 12 and 24, suggesting an annual season.

bricksq

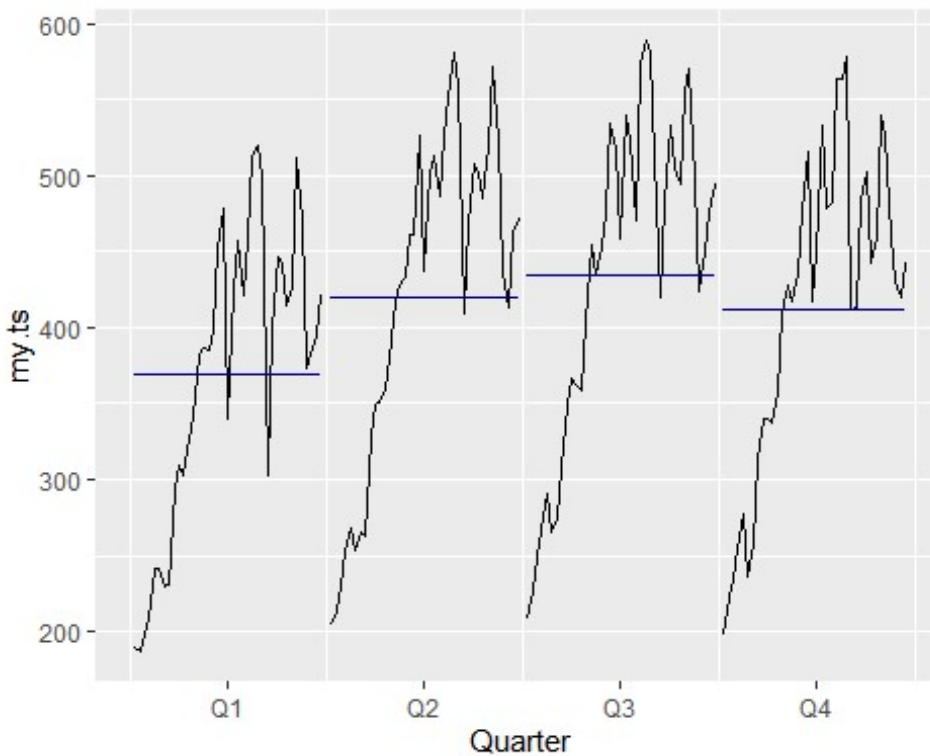
```
my.ts <- bricksq  
autoplot(my.ts)
```



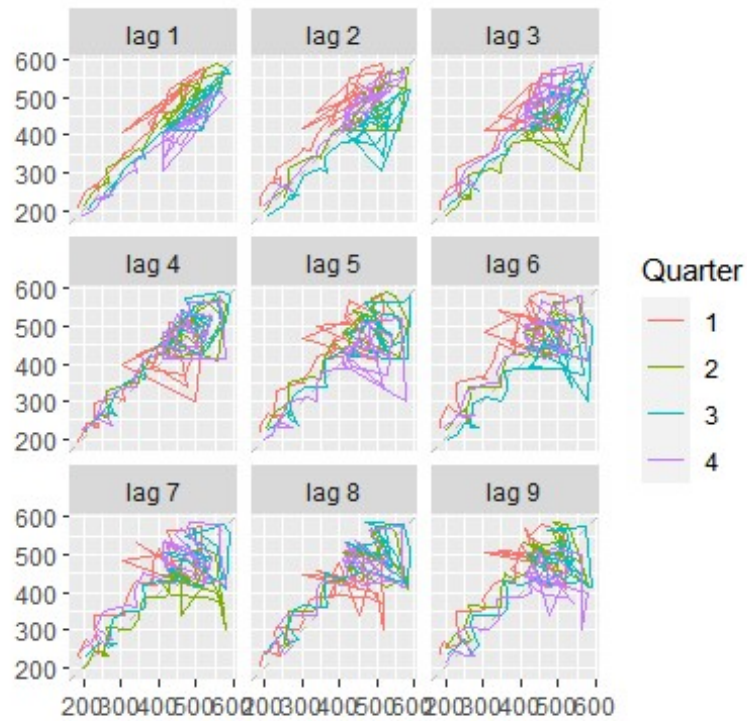
```
ggseasonplot(my.ts)
```

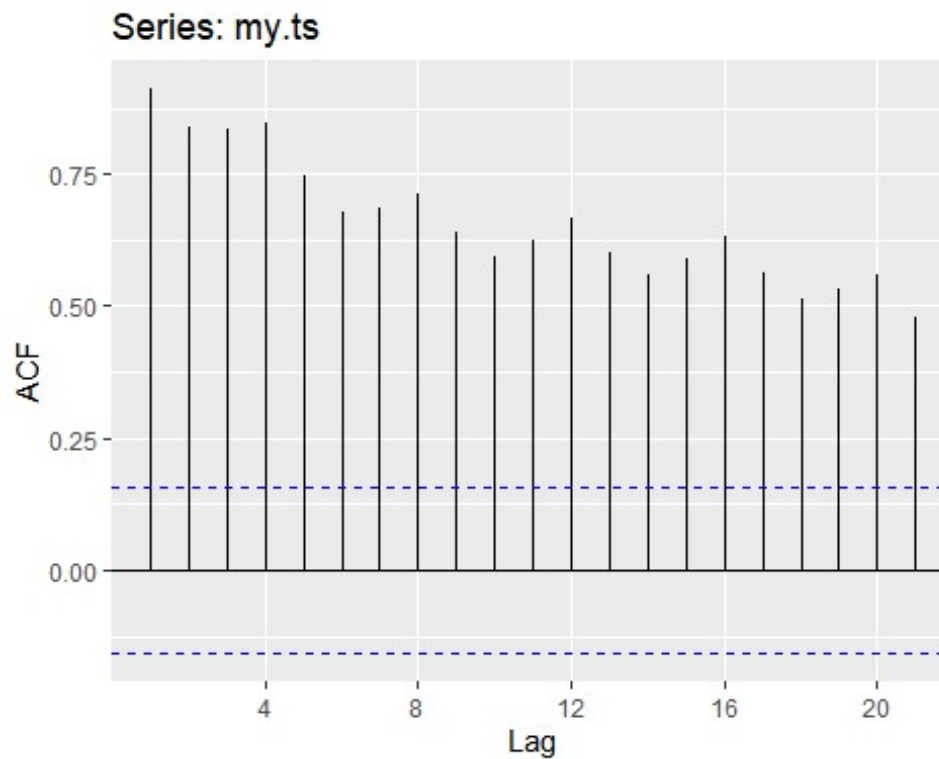
`ggsubseriesplot(my.ts)`



`gglagplot(my.ts)`



`ggAcf(my.ts)`

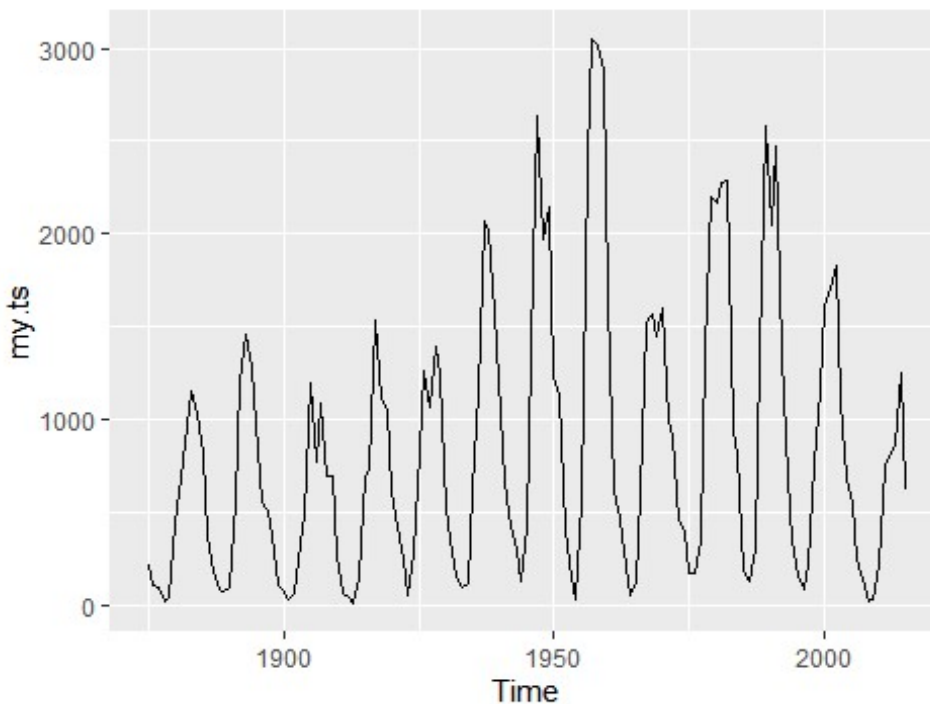


- This data shows an upward trend and a seemingly regular moves around that trend.

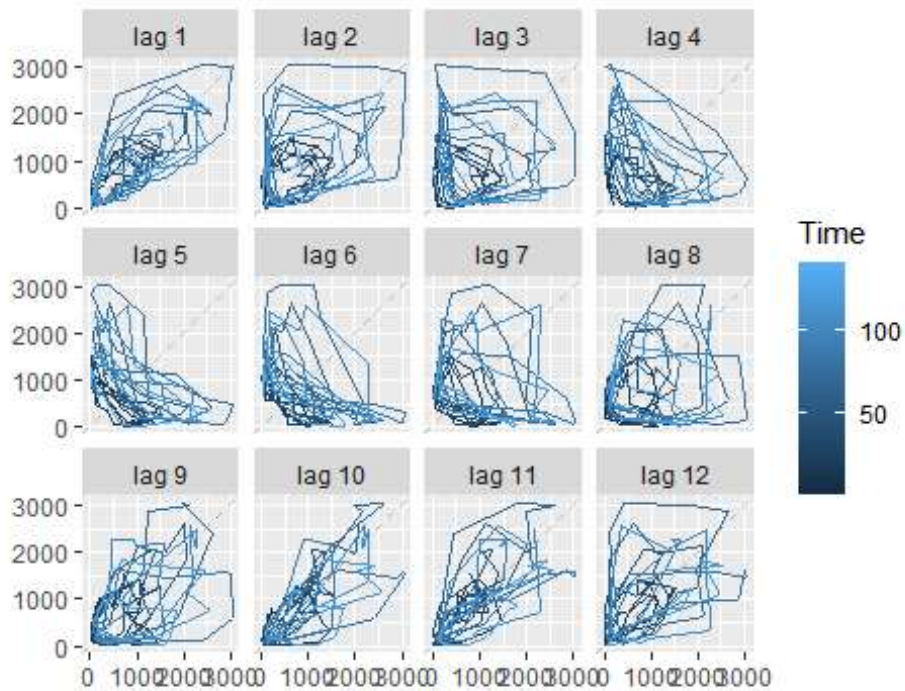
- The seasonality plots seem to indicate that Q1 is the lowest point in there year, otherwise the data appear to be somewhat flat.
- The lag plots show the highest degree of relationship at a lag of 1, suggesting serial dependence with a 1Q lag. I wonder what this data would look like with monthly data rather than quarterly...
- The ACF shows reasonably strong relationships with slow decay - information similar to what we say in the lag-plots.

sunspotarea

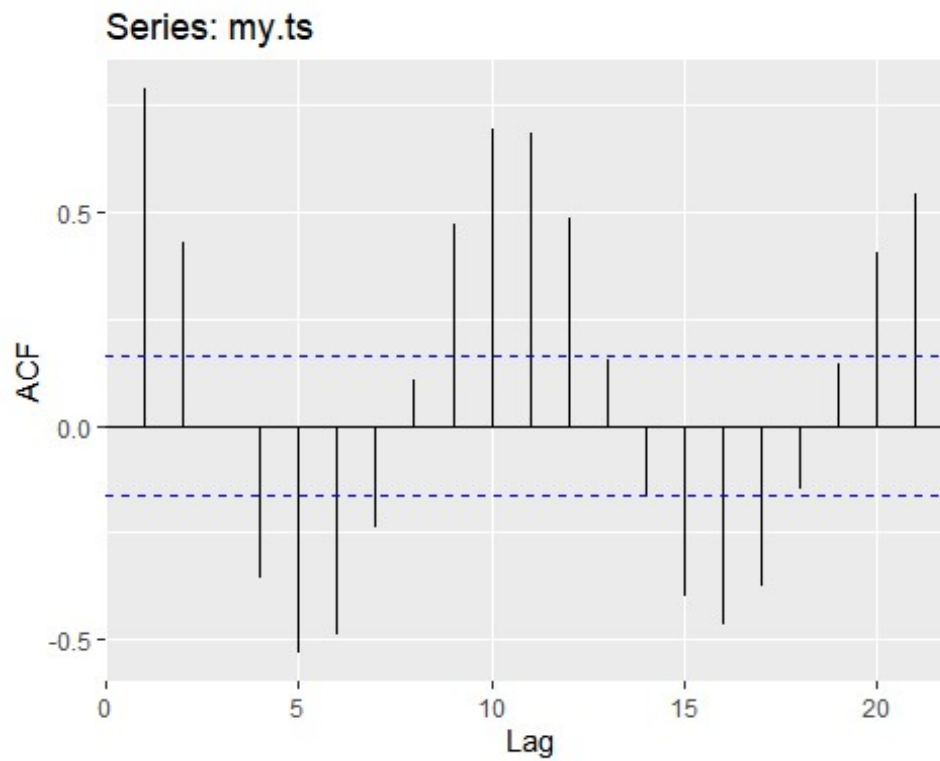
```
my.ts <- sunspotarea
autoplot(my.ts)
```



```
#ggseasonplot(my.ts)
#ggsubseriesplot(my.ts)
gglagplot(my.ts, lags=12)
```



`ggAcf(my.ts)`

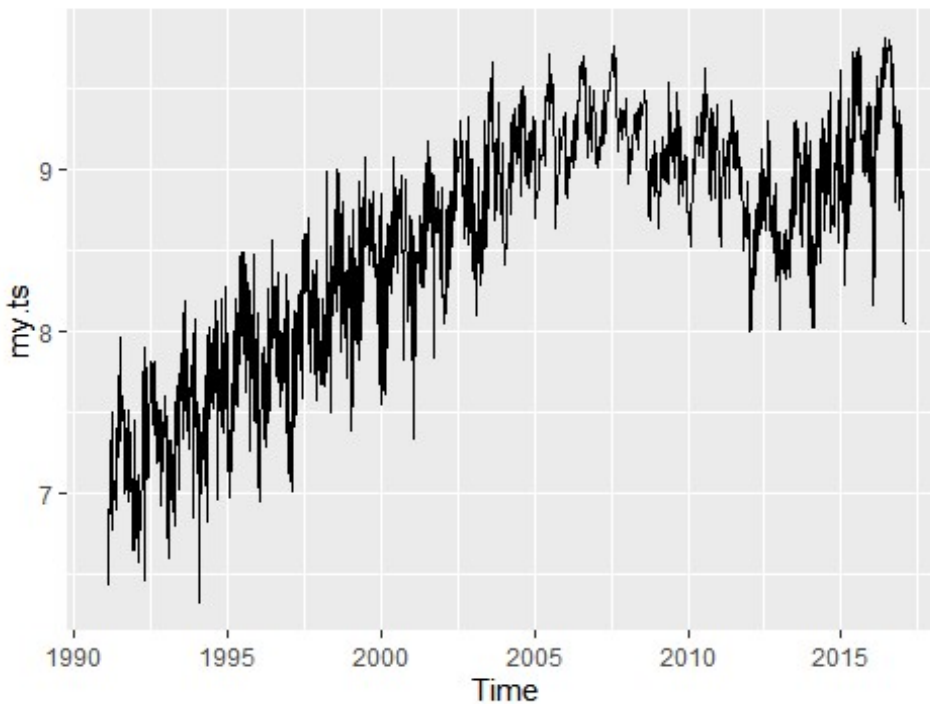


- This data shows a clear cycle and no trend.

- Given that the data is annual observations, there is no “seasonal” effect.
- The lag plots show the most dispersion at the bottom of the cycle (panel 3 - 6) and the least, or at least lower dispersion near the top of the cycle (panel 9 - 11)
- The ACF is also indicative of a cyclical pattern in that it looks like a sine wave with nadirs and peaks every 5 & 10 years.

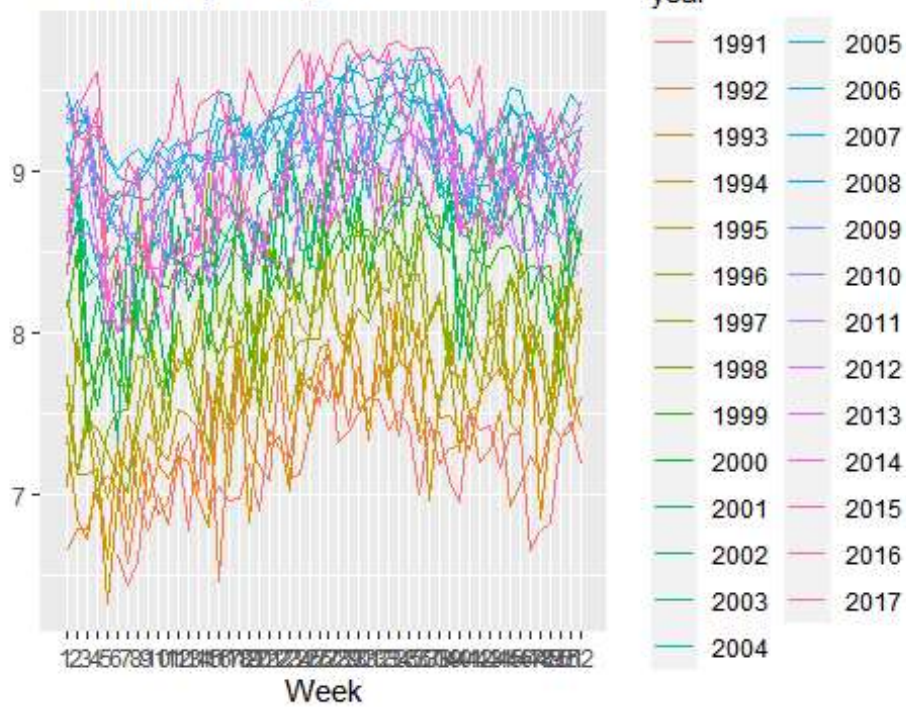
gasoline

```
my.ts <- gasoline  
autoplot(my.ts)
```



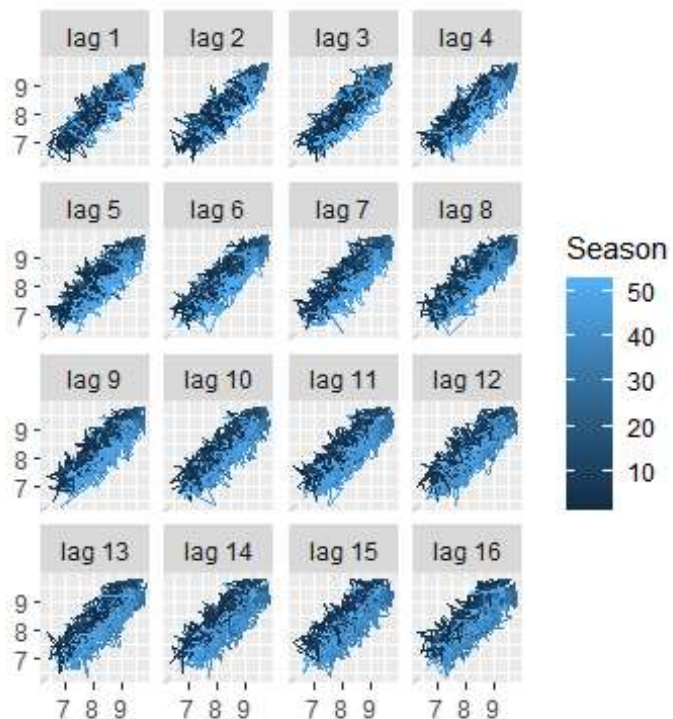
```
ggseasonplot(my.ts)
```


Seasonal plot: my.ts

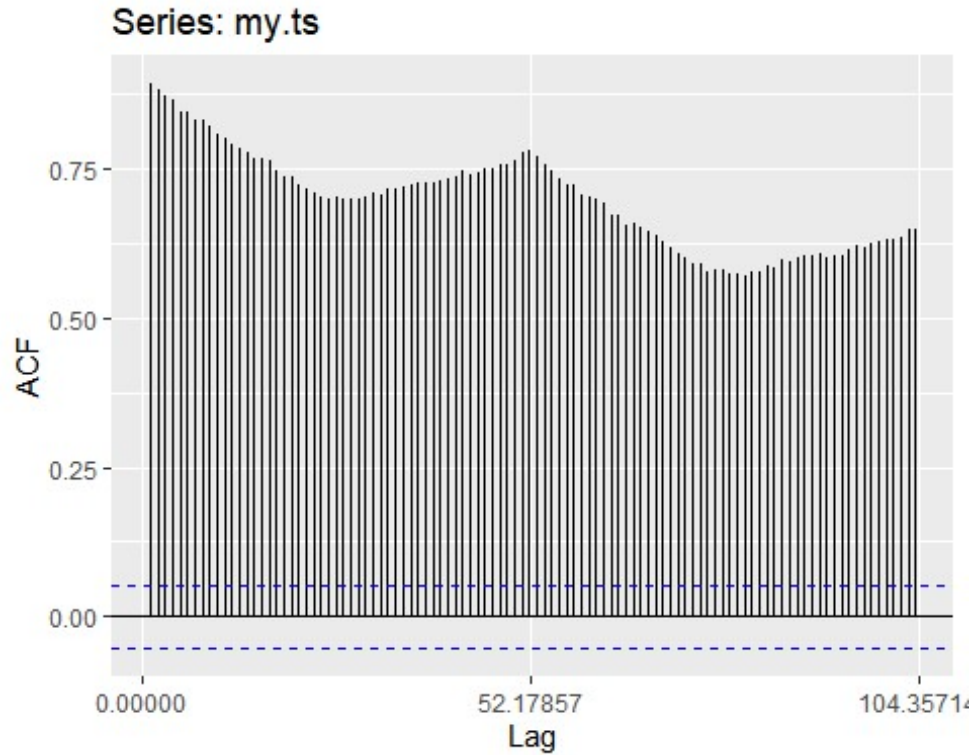


`#ggsubseriesplot(my.ts)`

`gglagplot(my.ts)`



```
ggAcf(my.ts)
```



- This data shows a noisy up trend with apparent seasonality.
- The seasonal plot shows a peak near weeks 0-4 and a lull for the few weeks following.
- The lag plots show a high degree of relatedness in all cases
- The ACF shows a slow and consistent decay possibly of an annual season.