Data624 - Homework1

Amit Kapoor

02/06/2021

```
library(fpp2)
```

Exercise

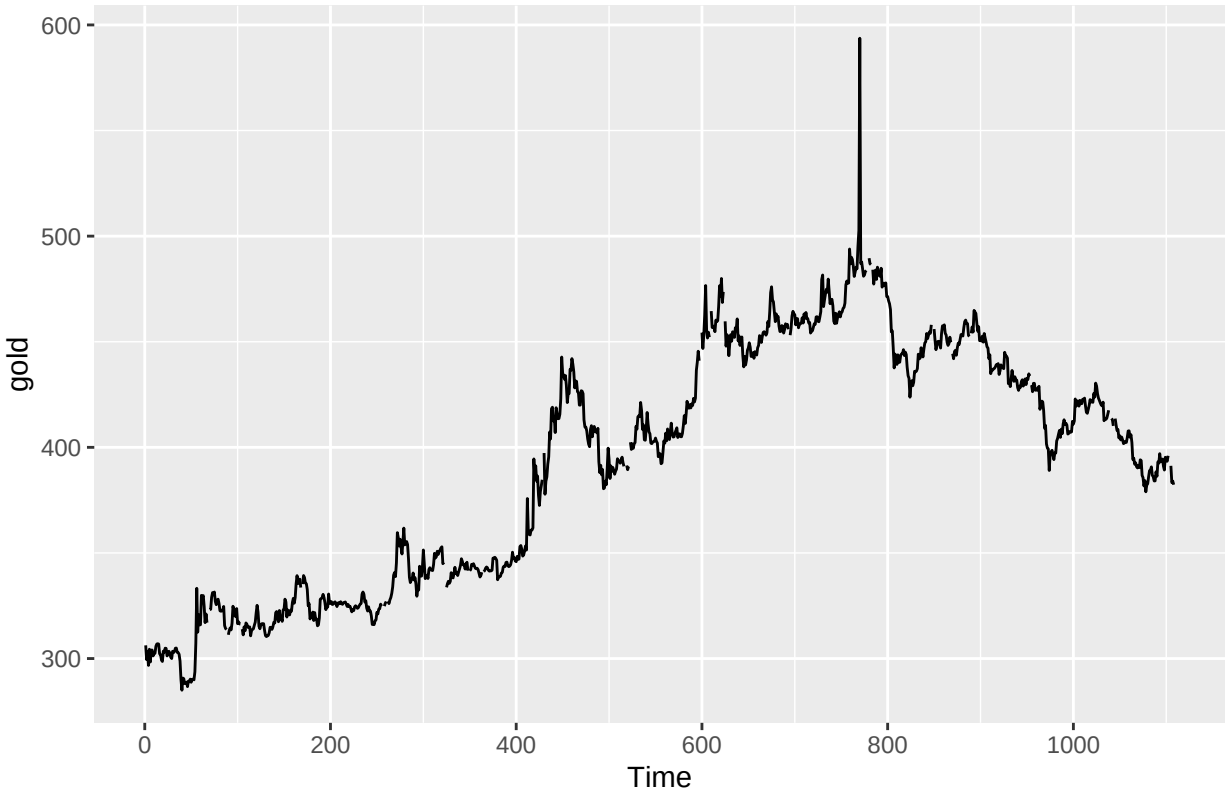
2.1 Use the help function to explore what the series gold, woolyrnq and gas represent.

```
?gold
?woolyrnq
?gas
```

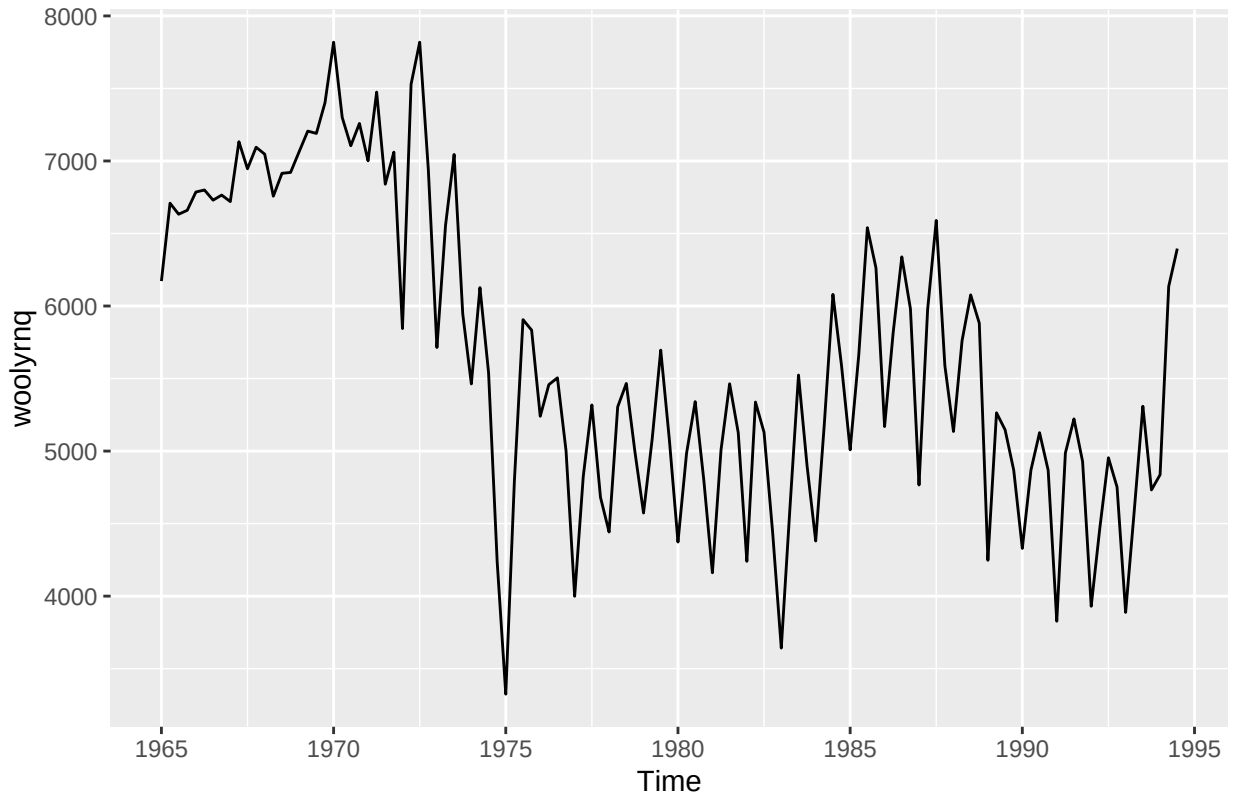
- gold: Daily morning gold prices in US dollars. 1 January 1985 – 31 March 1989.
- woolyrnq: Quarterly production of woollen yarn in Australia: tonnes. Mar 1965 – Sep 1994.
- gas: Australian monthly gas production: 1956–1995.

a. Use autoplot() to plot each of these in separate plots.

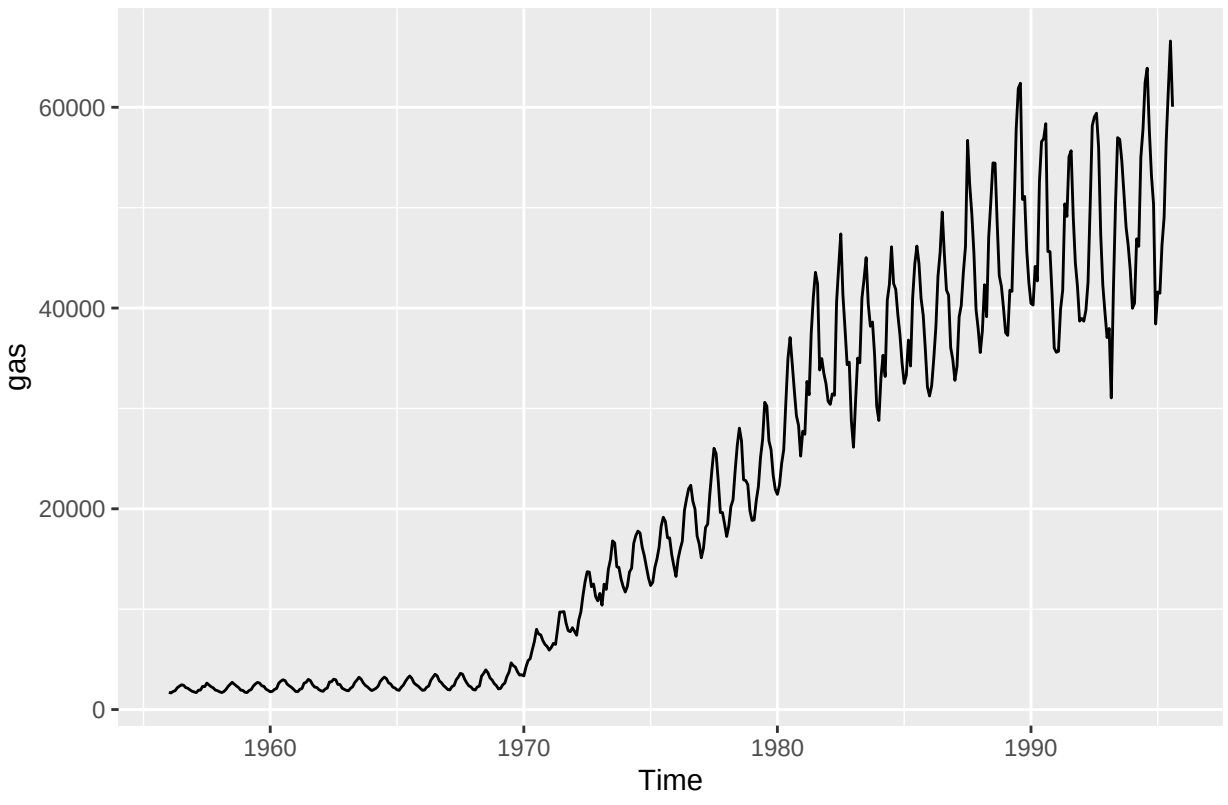
```
autoplot(gold)
```



```
autoplot(woolyrnq)
```




```
autoplot(gas)
```



b. What is the frequency of each series? Hint: apply the frequency() function.

```
frequency(gold)
```

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

```
frequency(woolyrnq)
```

```
## [1] 4
```

```
frequency(gas)
```

```
## [1] 12
```

gold, woolyrng and gas are yearly, quarterly and monthly time series respectively.

c. Use which.max() to spot the outlier in the gold series. Which observation was it?

```
which.max(gold)
```

```
## [1] 770
```

```
#outlier value
gold[which.max(gold)]
```

```
## [1] 593.7
```

It is 770 observation having gold price as 593.7.

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("https://otexts.com/fpp2/extrafiles/tute1.csv", header=TRUE)
head(tute1)
```

```
##        X  Sales AdBudget   GDP
## 1 Mar-81 1020.2    659.2 251.8
## 2 Jun-81  889.2    589.0 290.9
## 3 Sep-81  795.0    512.5 290.8
## 4 Dec-81 1003.9    614.1 292.4
## 5 Mar-82 1057.7    647.2 279.1
## 6 Jun-82  944.4    602.0 254.0
```

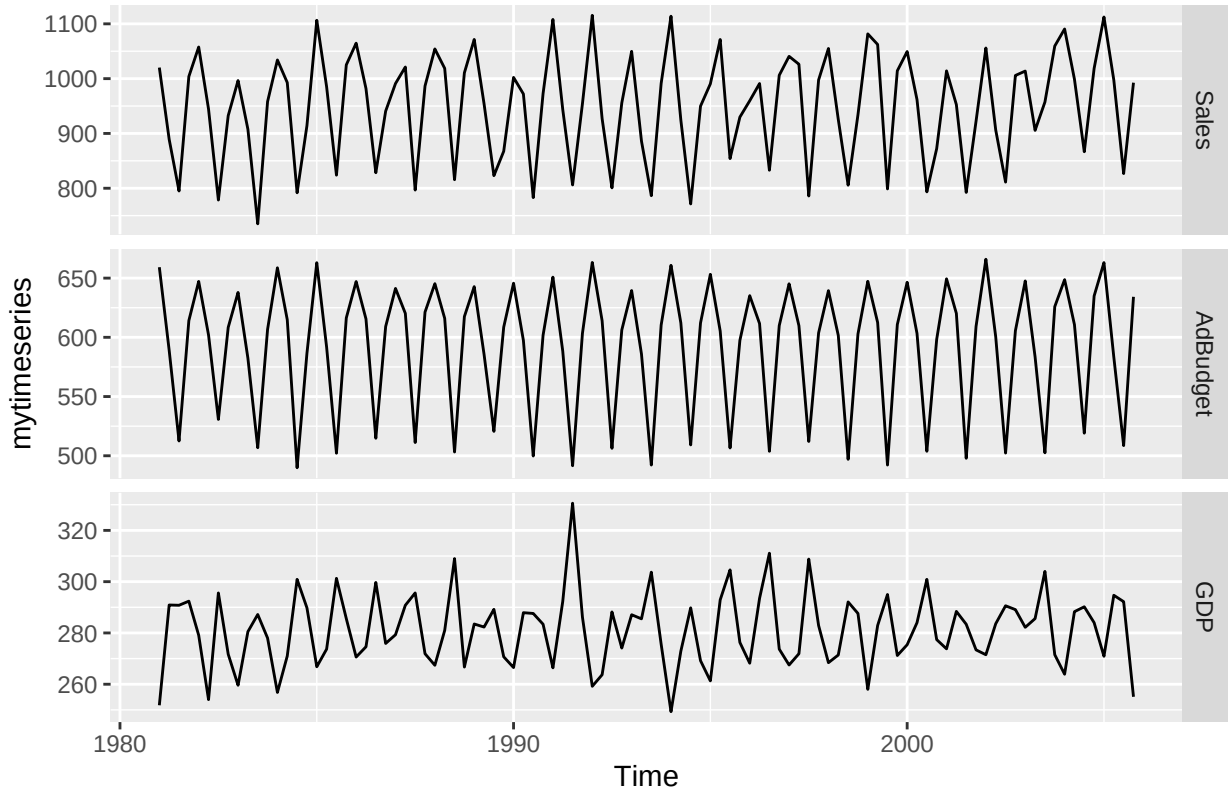
b. Convert the data to time series

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

```
##          Sales AdBudget   GDP
## 1981 Q1 1020.2    659.2 251.8
## 1981 Q2  889.2    589.0 290.9
## 1981 Q3  795.0    512.5 290.8
## 1981 Q4 1003.9    614.1 292.4
## 1982 Q1 1057.7    647.2 279.1
## 1982 Q2  944.4    602.0 254.0
```

c. Construct time series plots of each of the three series.

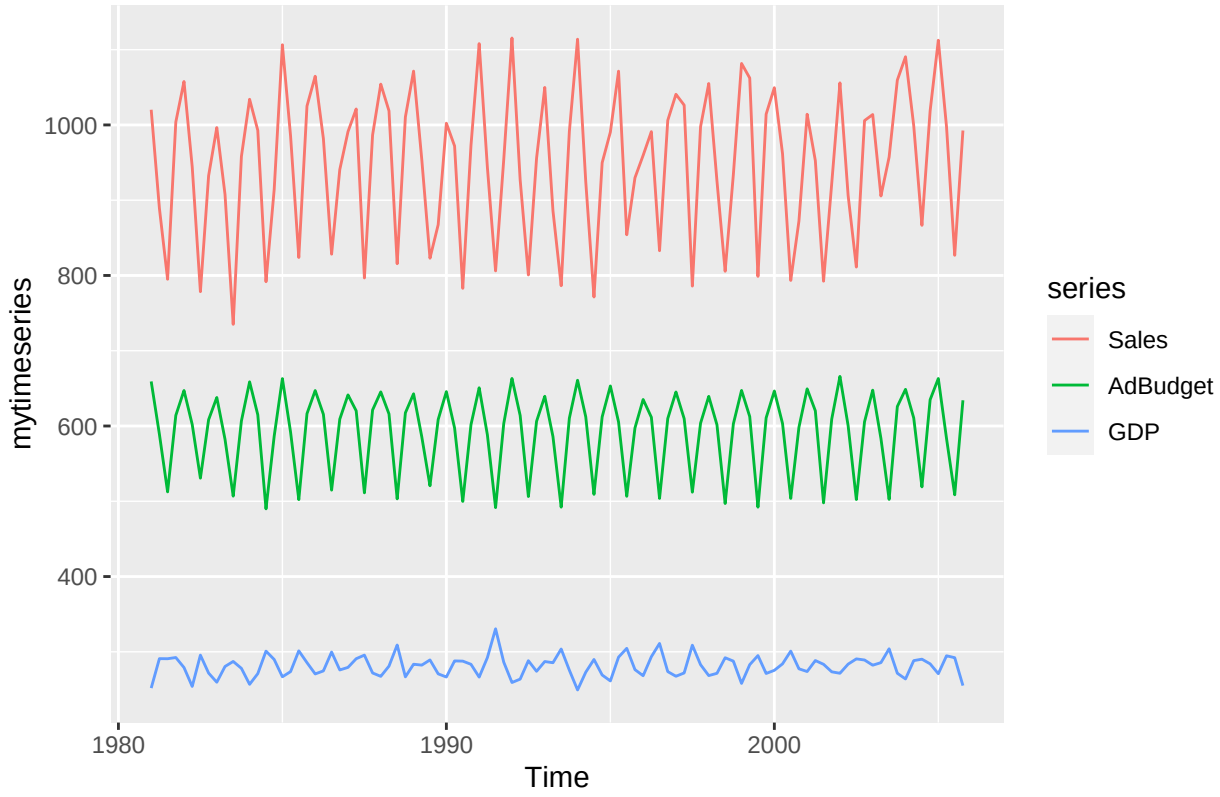
```
autoplot(mytimeseries, facets = TRUE)
```



Check what happens when you don't include facets=TRUE?

If we don't include facets=TRUE, it gets plotted on single axis and each series is assigned a color.

```
autoplot(mytimeseries)
```



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

a. You can read the data into R with the following script

```
retaildata <- readxl::read_excel("retail.xlsx", skip=1)
head(retaildata)
```

```
## # A tibble: 6 x 190
##   `Series ID`         A3349335T A3349627V A3349338X A3349398A A3349468W
##   <dttm>                  <dbl>     <dbl>     <dbl>     <dbl>     <dbl>
## 1 1982-04-01 00:00:00      303.      41.7      63.9      409.      65.8
## 2 1982-05-01 00:00:00      298.      43.1      64        405.      65.8
## 3 1982-06-01 00:00:00      298       40.3      62.7      401       62.3
## 4 1982-07-01 00:00:00      308.      40.9      65.6      414.      68.2
## 5 1982-08-01 00:00:00      299.      42.1      62.6      404.      66
## 6 1982-09-01 00:00:00      305.      42        64.4      412.      62.3
## # ... with 184 more variables: A3349336V <dbl>, A3349337W <dbl>,
## #   A3349397X <dbl>, A3349399C <dbl>, A3349874C <dbl>, A3349871W <dbl>,
## #   A3349790V <dbl>, A3349556W <dbl>, A3349791W <dbl>, A3349401C <dbl>,
## #   A3349873A <dbl>, A3349872X <dbl>, A3349709X <dbl>, A3349792X <dbl>,
## #   A3349789K <dbl>, A3349555V <dbl>, A3349565X <dbl>, A3349414R <dbl>,
## #   A3349799R <dbl>, A3349642T <dbl>, A3349413L <dbl>, A3349564W <dbl>,
## #   A3349416V <dbl>, A3349643V <dbl>, A3349483V <dbl>, A3349722T <dbl>,
## #   A3349727C <dbl>, A3349641R <dbl>, A3349639C <dbl>, A3349415T <dbl>,
## #   A3349349F <dbl>, A3349563V <dbl>, A3349350R <dbl>, A3349640L <dbl>,
## #   A3349566A <dbl>, A3349417W <dbl>, A3349352V <dbl>, A3349882C <dbl>,
## #   A3349561R <dbl>, A3349883F <dbl>, A3349721R <dbl>, A3349478A <dbl>,
## #   A3349637X <dbl>, A3349479C <dbl>, A3349797K <dbl>, A3349477X <dbl>,
```

```
## #   A3349719C <dbl>, A3349884J <dbl>, A3349562T <dbl>, A3349348C <dbl>,
## #   A3349480L <dbl>, A3349476W <dbl>, A3349881A <dbl>, A3349410F <dbl>,
## #   A3349481R <dbl>, A3349718A <dbl>, A3349411J <dbl>, A3349638A <dbl>,
## #   A3349654A <dbl>, A3349499L <dbl>, A3349902A <dbl>, A3349432V <dbl>,
## #   A3349656F <dbl>, A3349361W <dbl>, A3349501L <dbl>, A3349503T <dbl>,
## #   A3349360V <dbl>, A3349903C <dbl>, A3349905J <dbl>, A3349658K <dbl>,
## #   A3349575C <dbl>, A3349428C <dbl>, A3349500K <dbl>, A3349577J <dbl>,
## #   A3349433W <dbl>, A3349576F <dbl>, A3349574A <dbl>, A3349816F <dbl>,
## #   A3349815C <dbl>, A3349744F <dbl>, A3349823C <dbl>, A3349508C <dbl>,
## #   A3349742A <dbl>, A3349661X <dbl>, A3349660W <dbl>, A3349909T <dbl>,
## #   A3349824F <dbl>, A3349507A <dbl>, A3349580W <dbl>, A3349825J <dbl>,
## #   A3349434X <dbl>, A3349822A <dbl>, A3349821X <dbl>, A3349581X <dbl>,
## #   A3349908R <dbl>, A3349743C <dbl>, A3349910A <dbl>, A3349435A <dbl>,
## #   A3349365F <dbl>, A3349746K <dbl>, ...
```

b. Select one of the time series as follows (but replace the column name with your own chosen column)

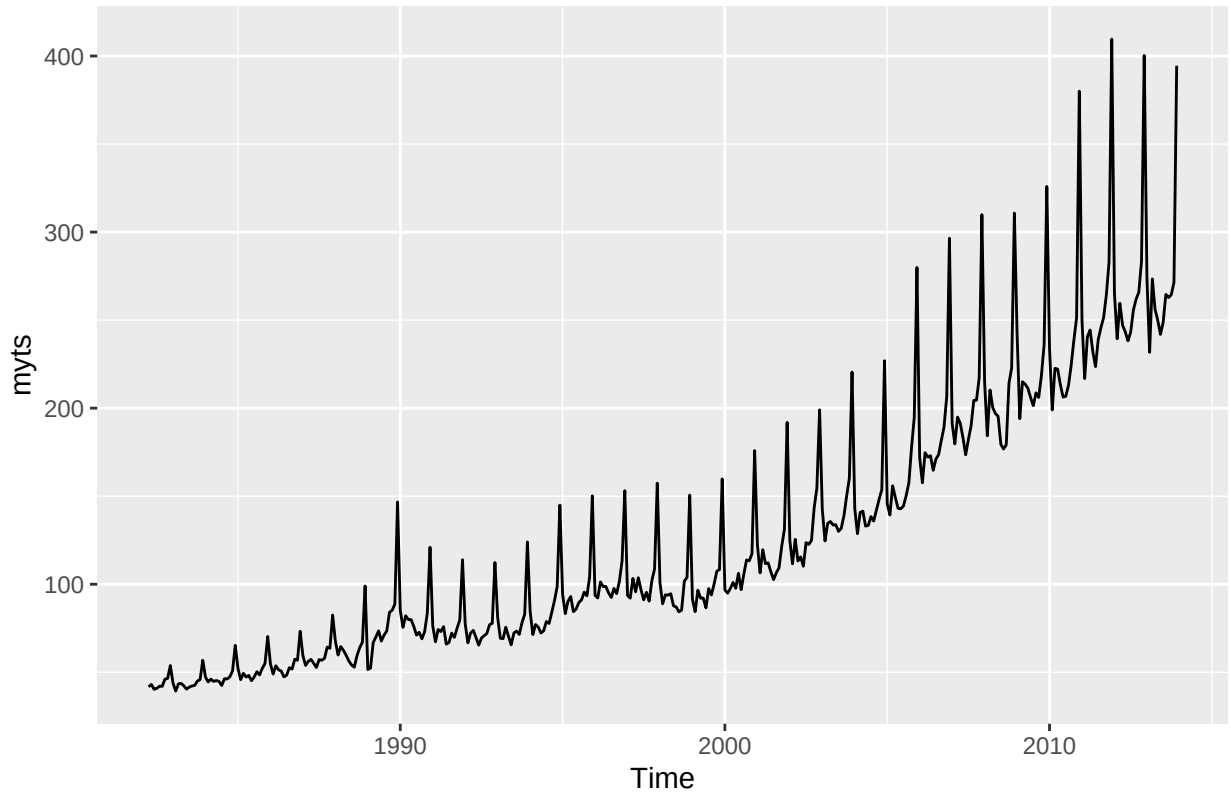
```
myts <- ts(retaildata[,"A3349627V"], frequency=12, start=c(1982,4))
head(myts)
```

```
##       Apr  May  Jun  Jul  Aug  Sep
## 1982 41.7 43.1 40.3 40.9 42.1 42.0
```

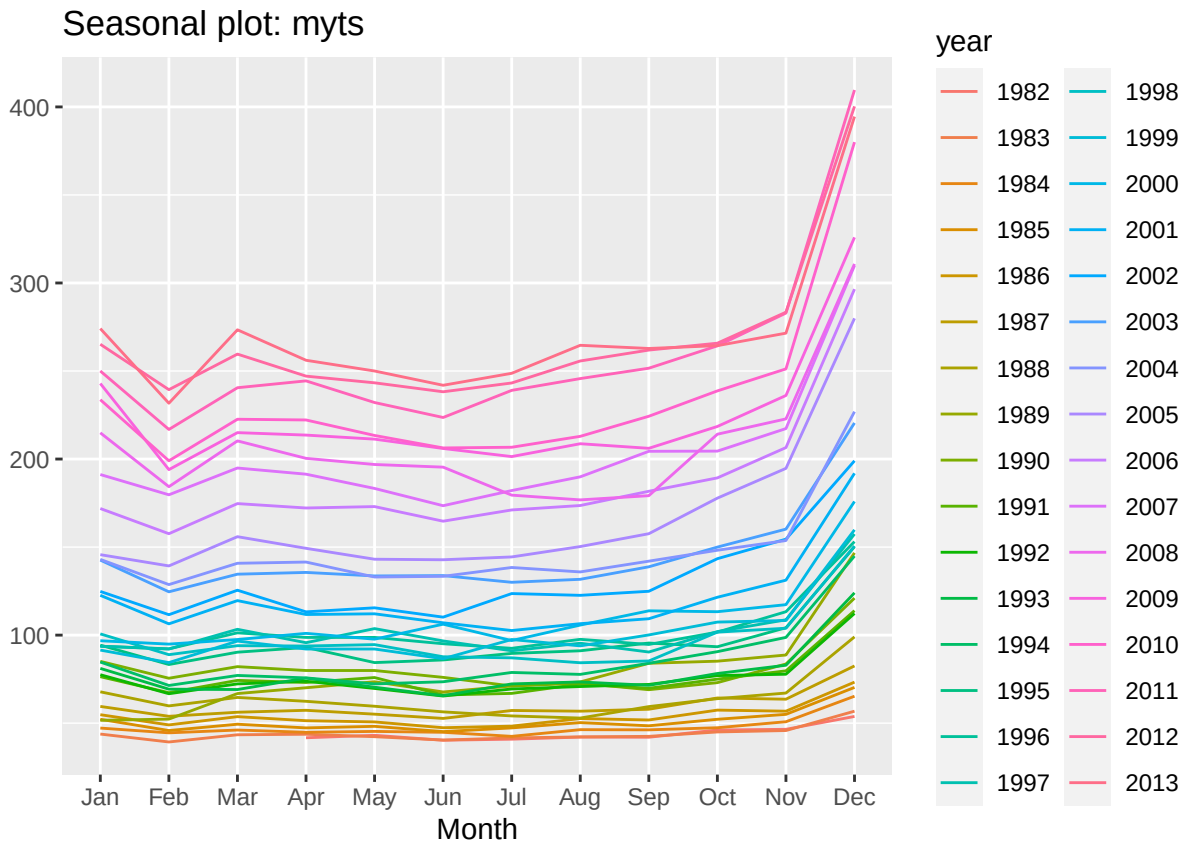
c. Explore your chosen retail time series using the following functions:

autoplot(), ggseasonplot(), ggsubseriesplot(), gglagplot(), ggAcf()

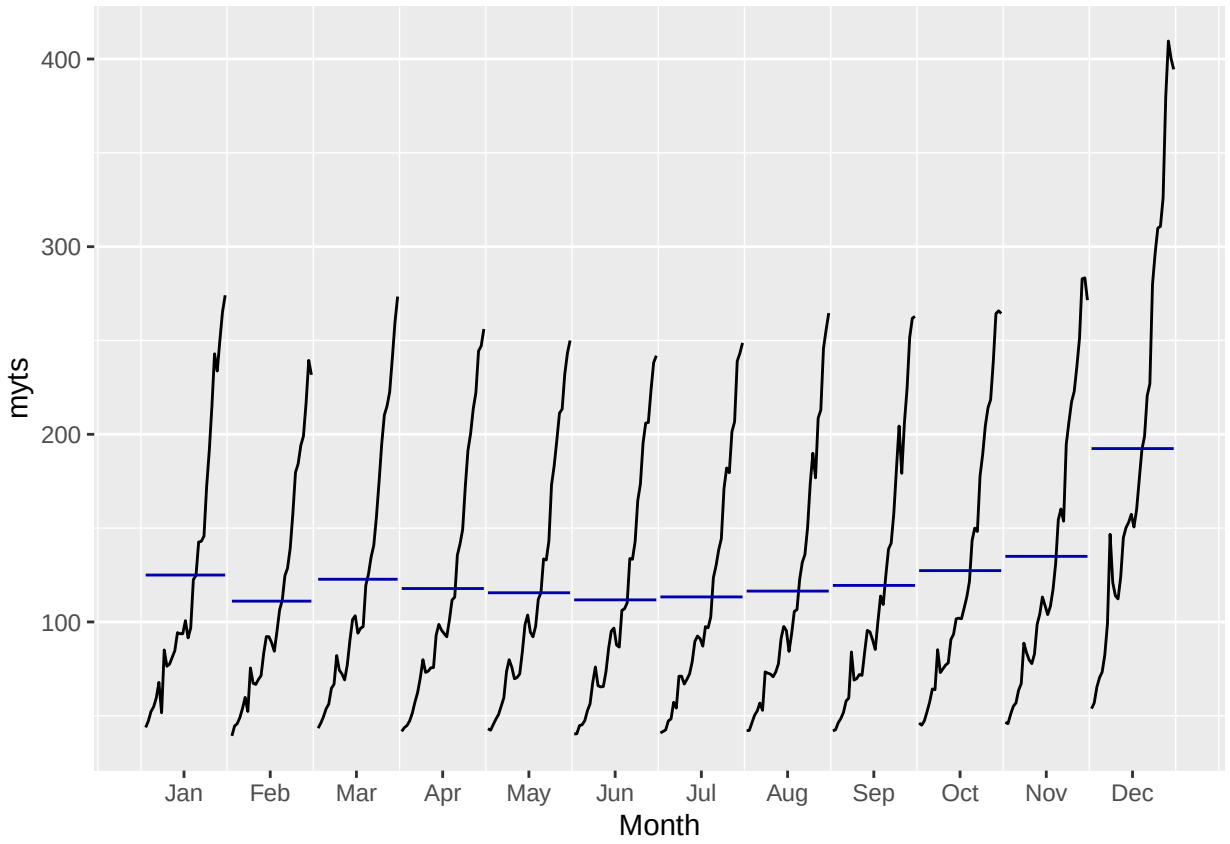
```
autoplot(myts)
```



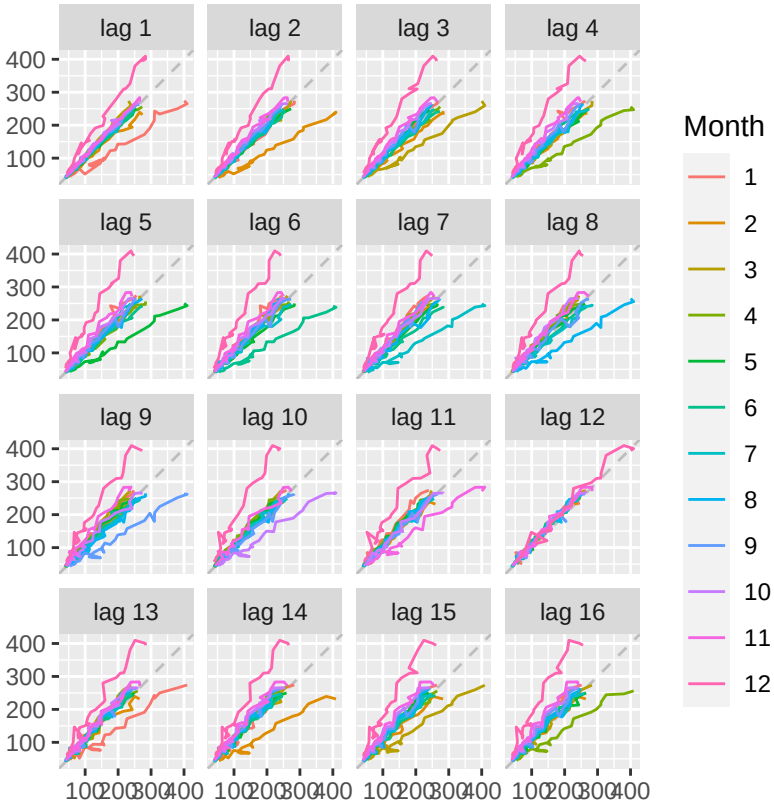

```
ggseasonplot(myts)
```



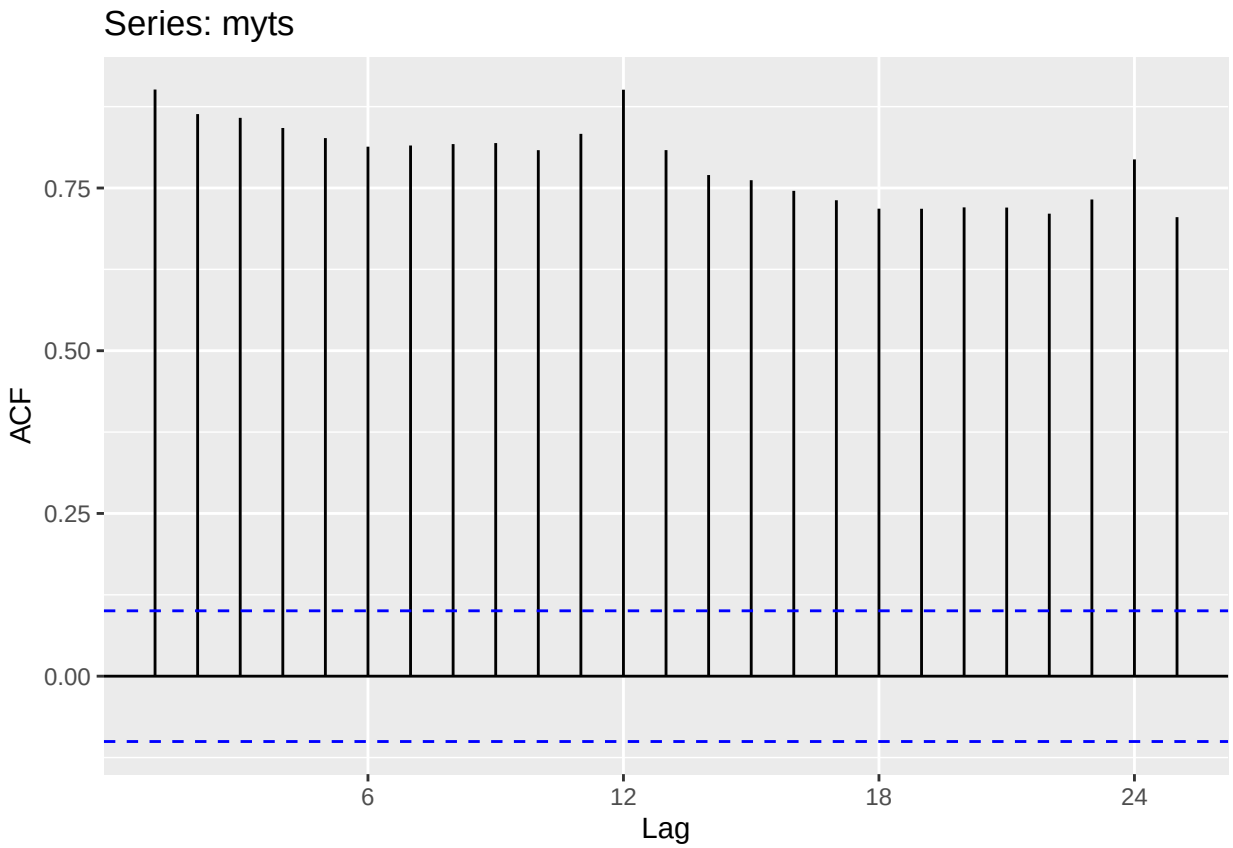

```
ggsubseriesplot(myts)
```




```
gglagplot(myts)
```



```
ggAcf(myts)
```



- Can you spot any seasonality, cyclicity and trend?

There is a clear annual seasonality increase in retail sales from October to December. I see a consistent upward trend and dont see cyclicity.

- What do you learn about the series?

There is an increase in retail season is the Christmas shopping season. The trend seems rising until the 1990 where it gets flattened out for alomost a decade. After 2000 the trend continues to go up.

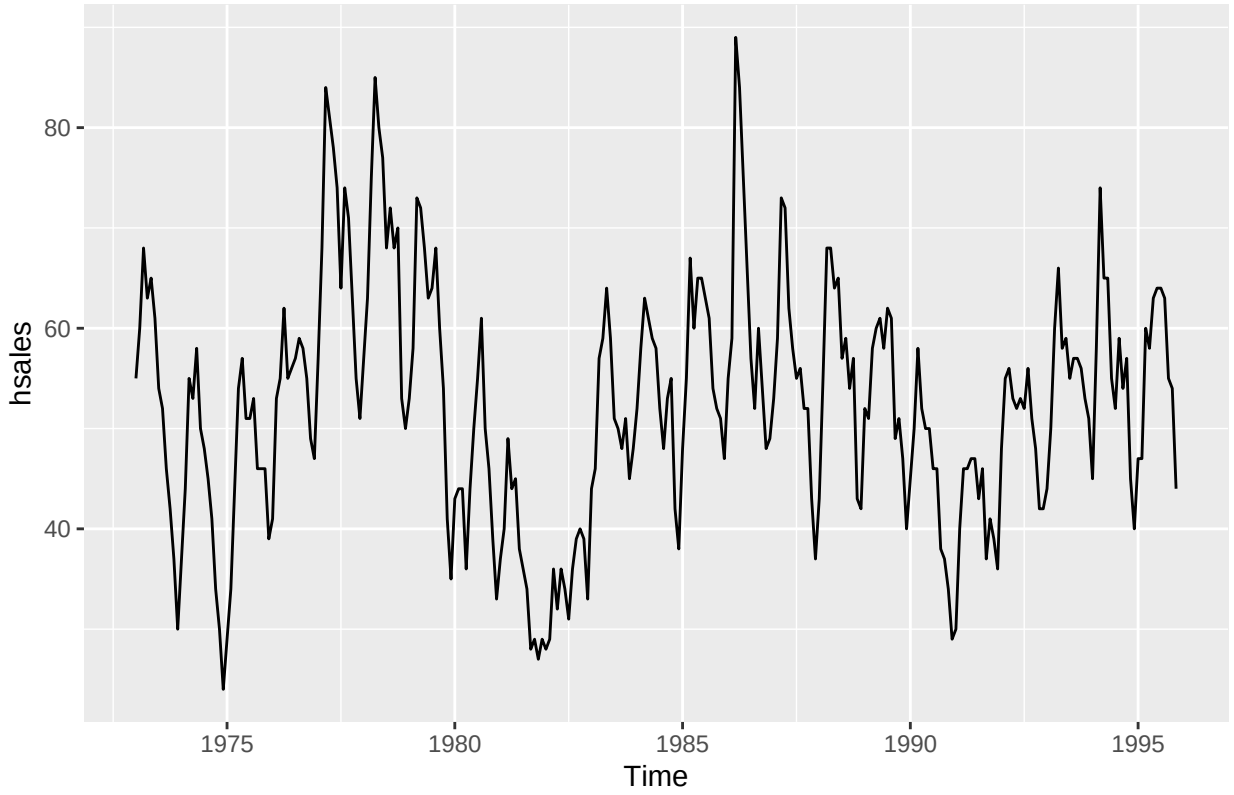
2.6 Use the following graphics functions: autoplot(), ggseasonplot(), ggsubseriesplot(), gglagplot(), ggAcf() and explore features from the following time series: hsales, usdeaths, bricksq, sunspotarea, gasoline.

Time series hsales

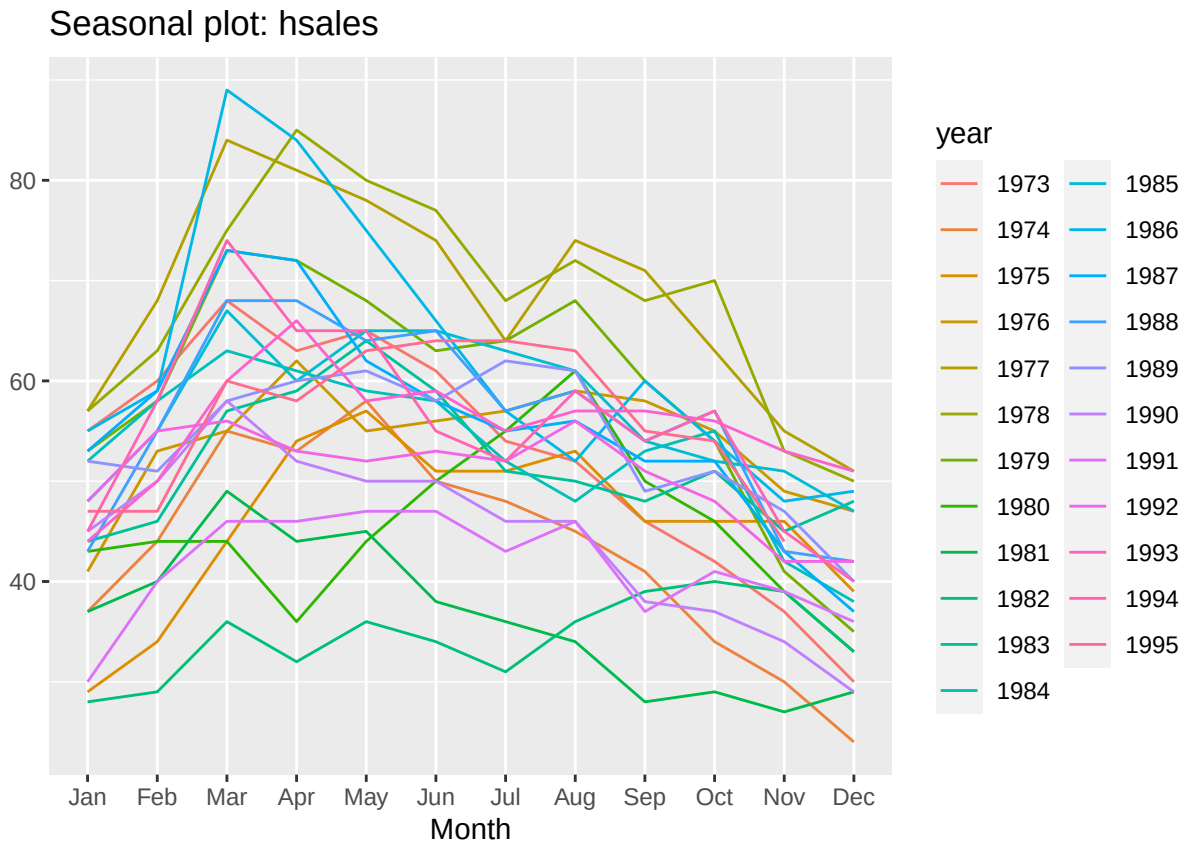
hsales - Monthly sales of new one-family houses sold in the USA since 1973.

```
?hsales
```

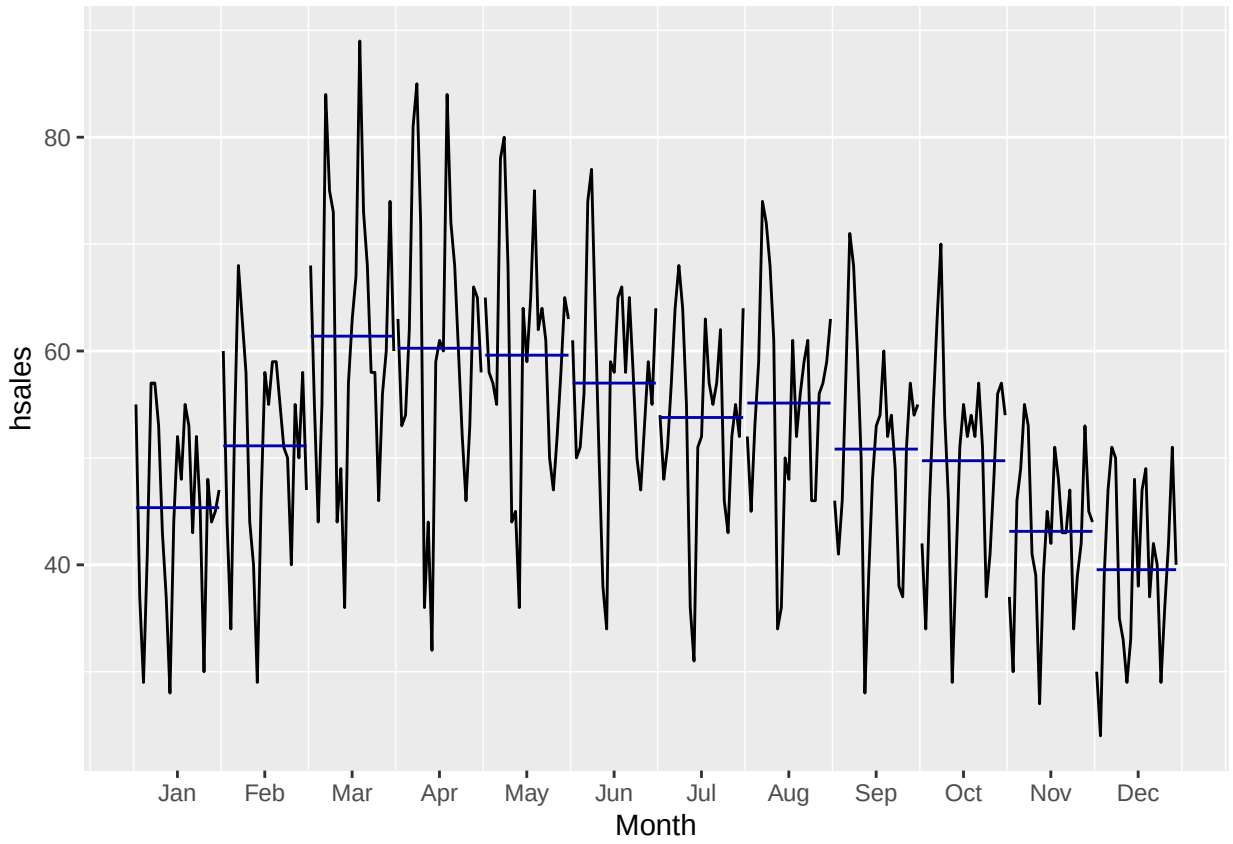
```
autoplot(hsales)
```



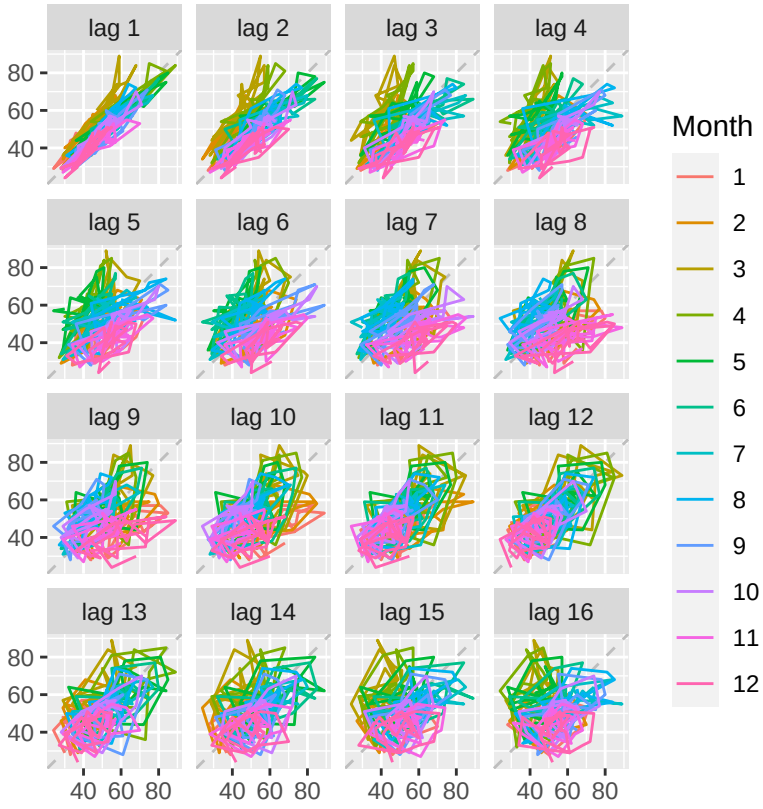

```
ggseasonplot(hsales)
```



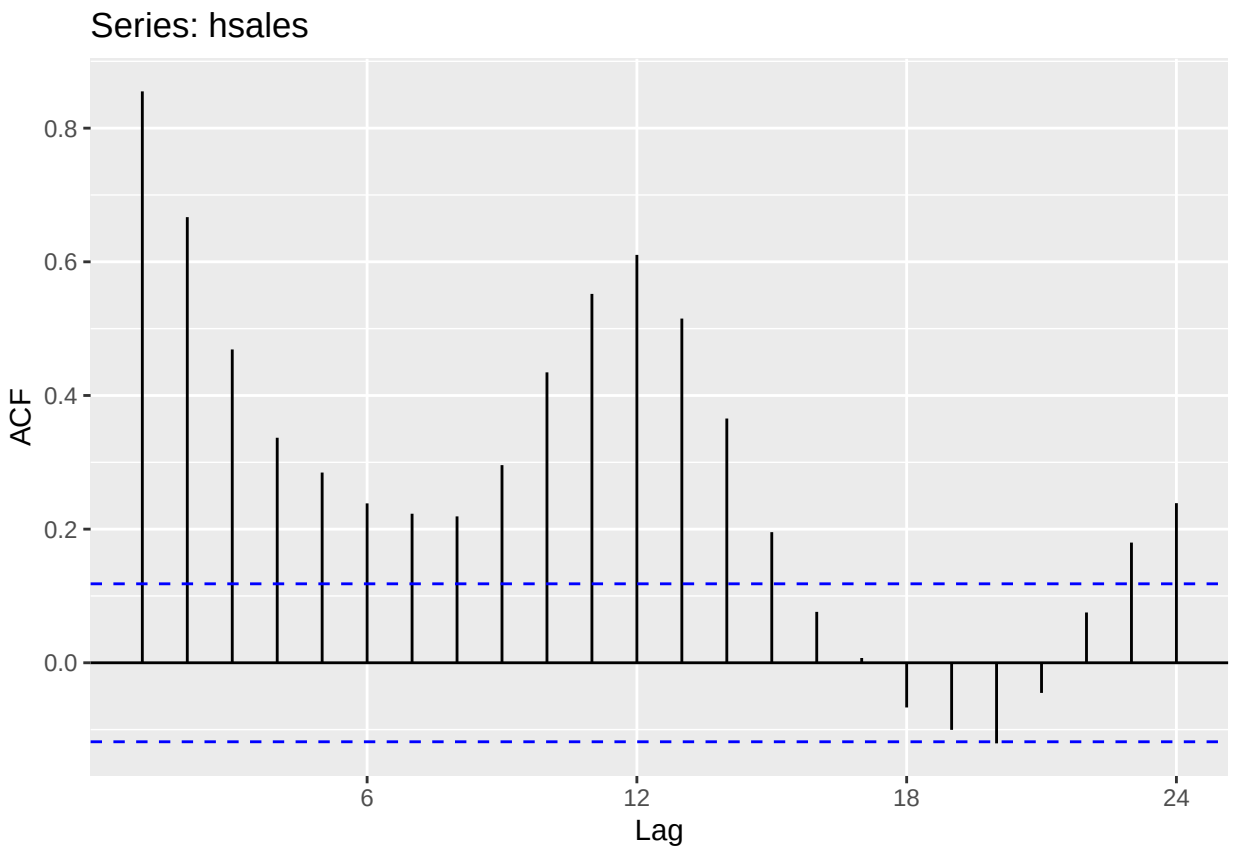

```
ggsubseriesplot(hsales)
```




```
gglagplot(hsales)
```



```
ggAcf(hsales)
```



a. Can you spot any seasonality, cyclicity and trend?

Seasonal and subseries plot shows the peak of sales of new one-family houses in March and trough in December so there is a seasonality in data. In approx every 10 years, I see a decrease in home sales. The lagplot shows strong linear relationship at lag 1 which keep on getting weak in upcoming lags until lag 12, that again shows the seasonality in data.

b. What do you learn about the series?

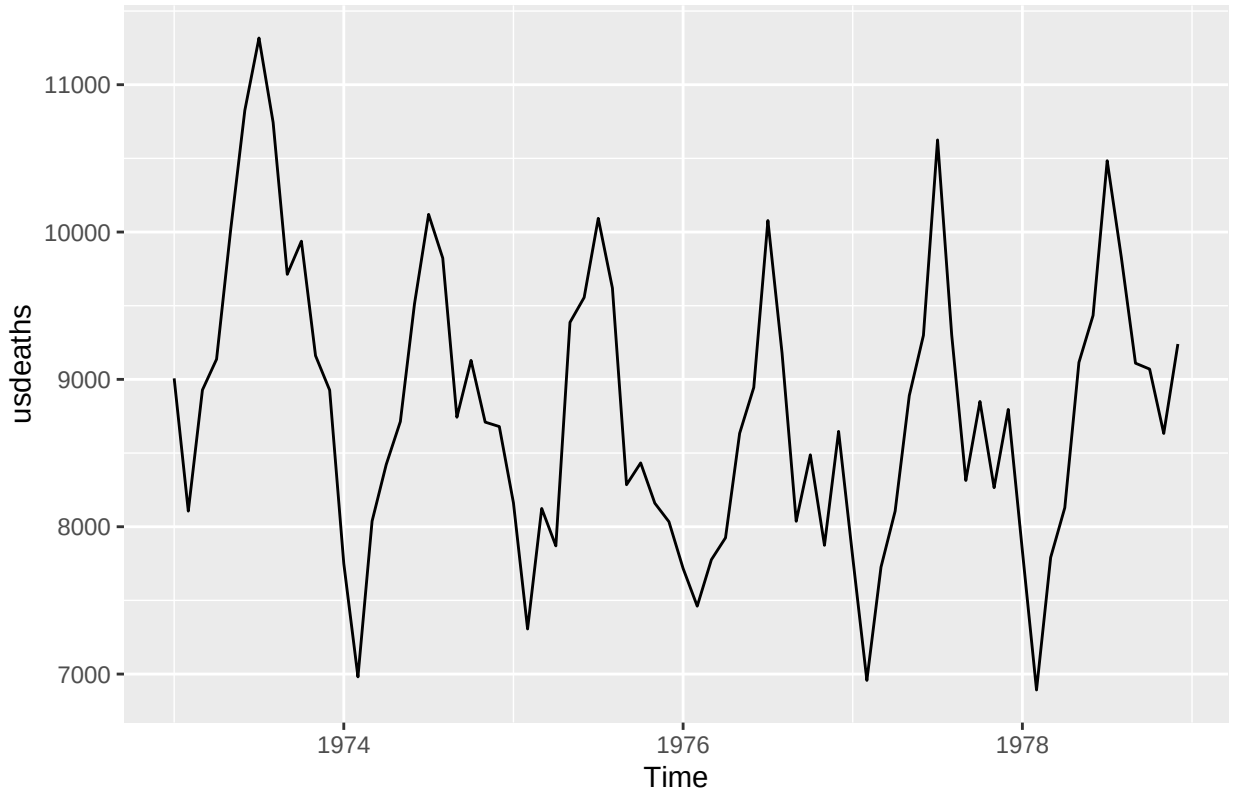
Realtors are going to be busy through Spring (March to May) as the home sale is highest during this time.

usdeaths

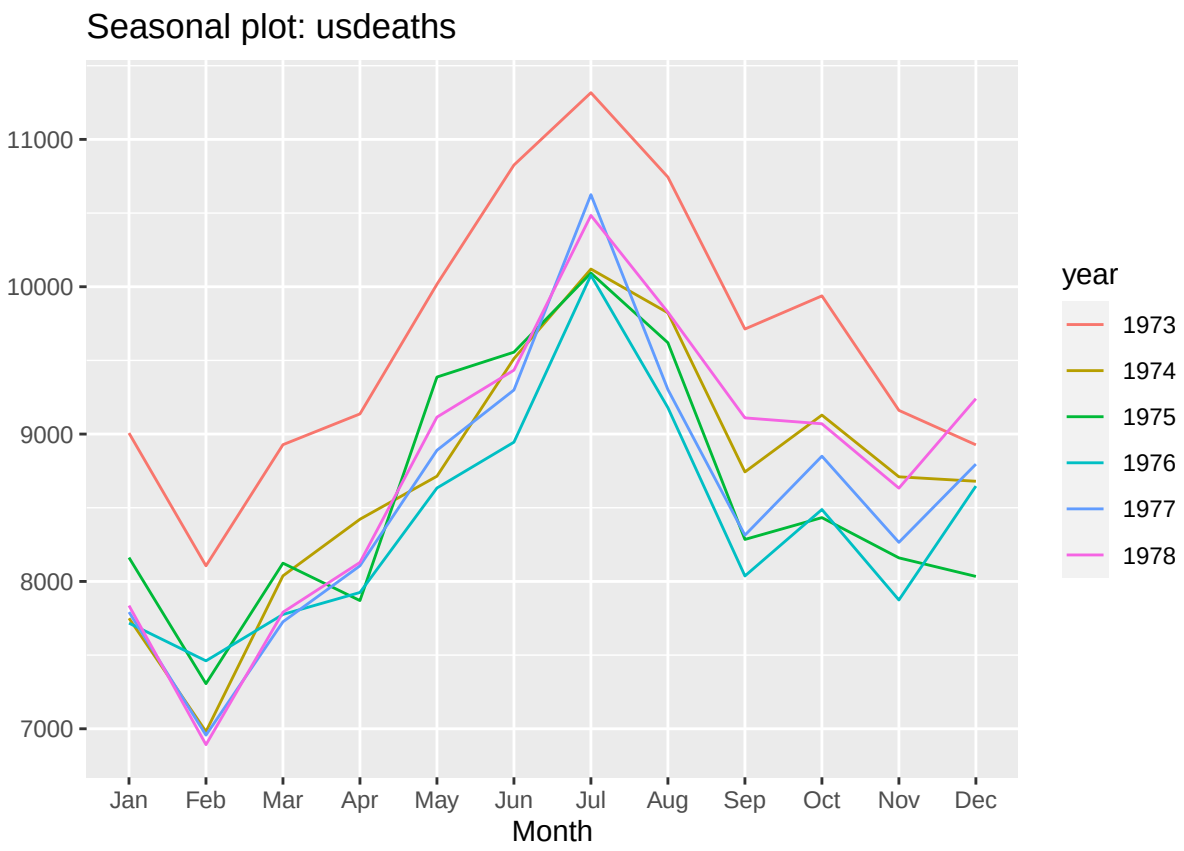
usdeaths - Monthly accidental deaths in USA.

```
?usdeaths
```

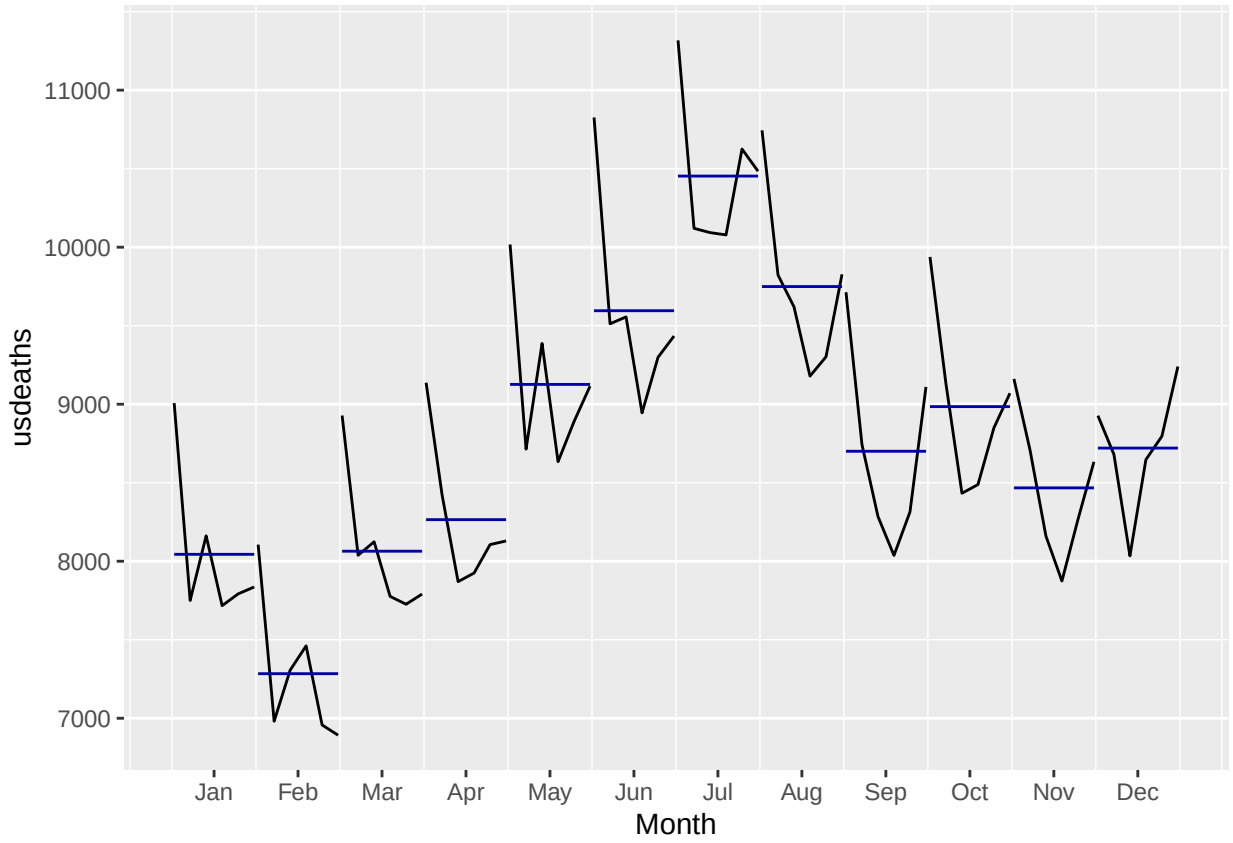
```
autoplot(usdeaths)
```



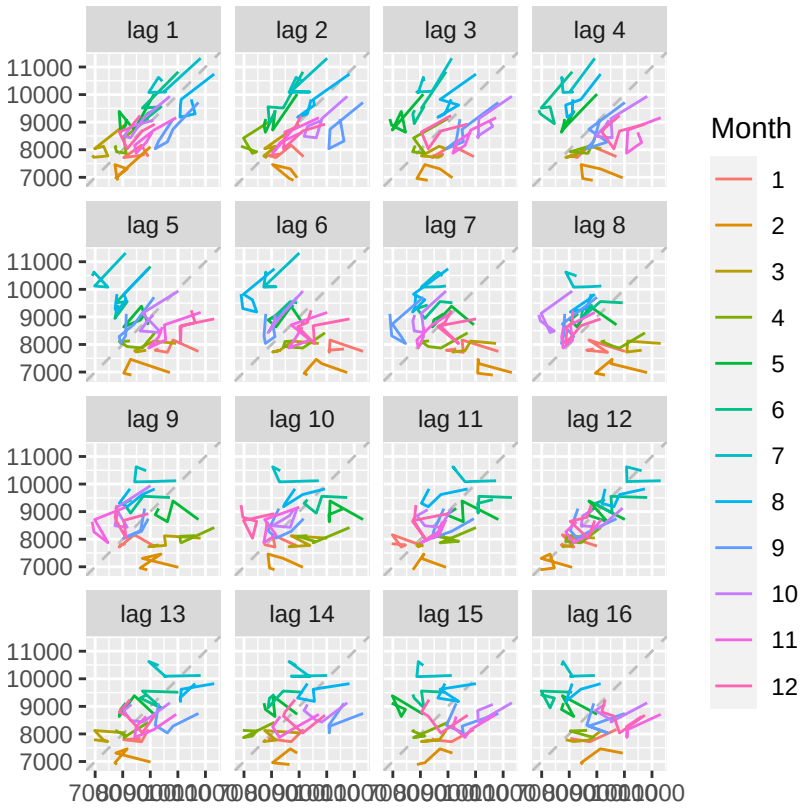

```
ggseasonplot(usdeaths)
```



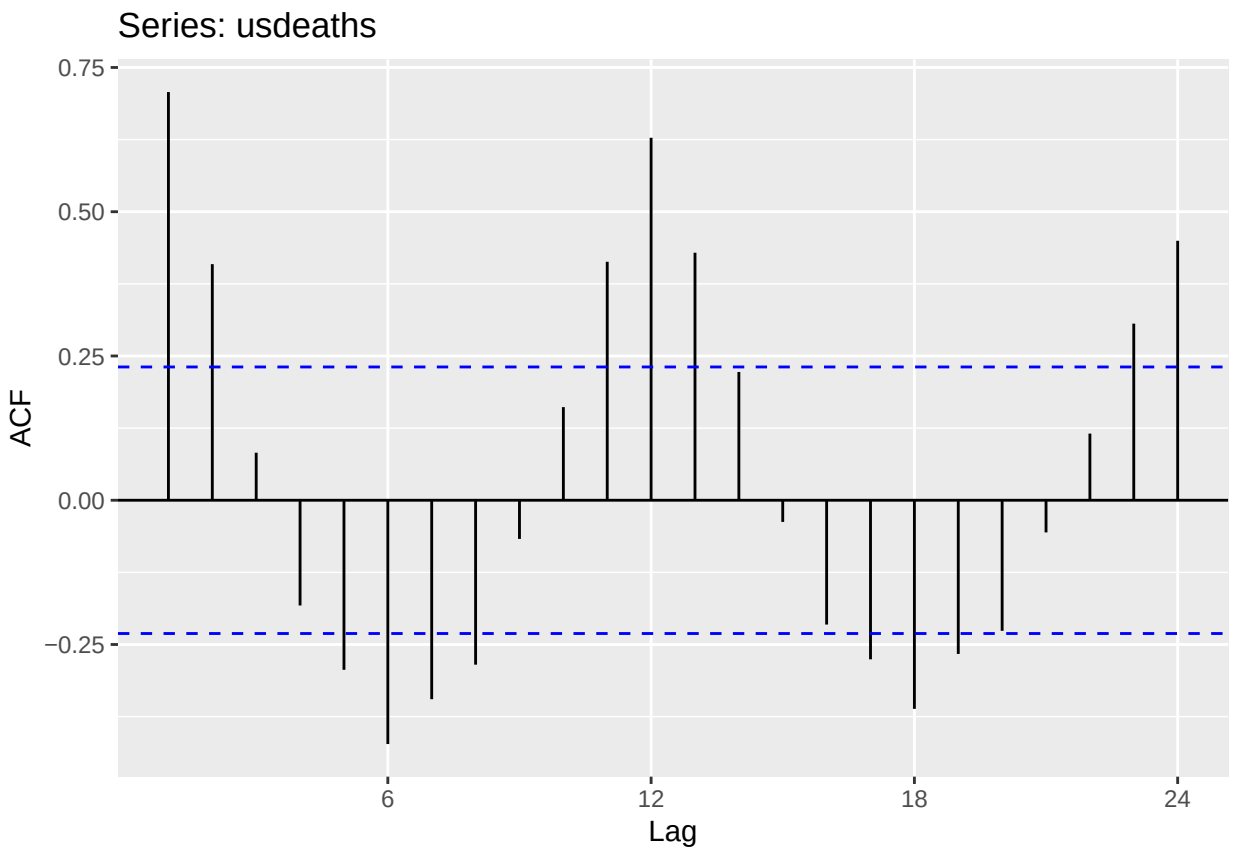
```
ggsubseriesplot(usdeaths)
```




```
gglagplot(usdeaths)
```




```
ggAcf(usdeaths)
```



a. Can you spot any seasonality, cyclicity and trend?

I see annual seasonality in usdeaths data having peak in July and trough in Feb. No trend or cyclicity observed. The lag plot 12 has strongest linear relationship.

b. What do you learn about the series?

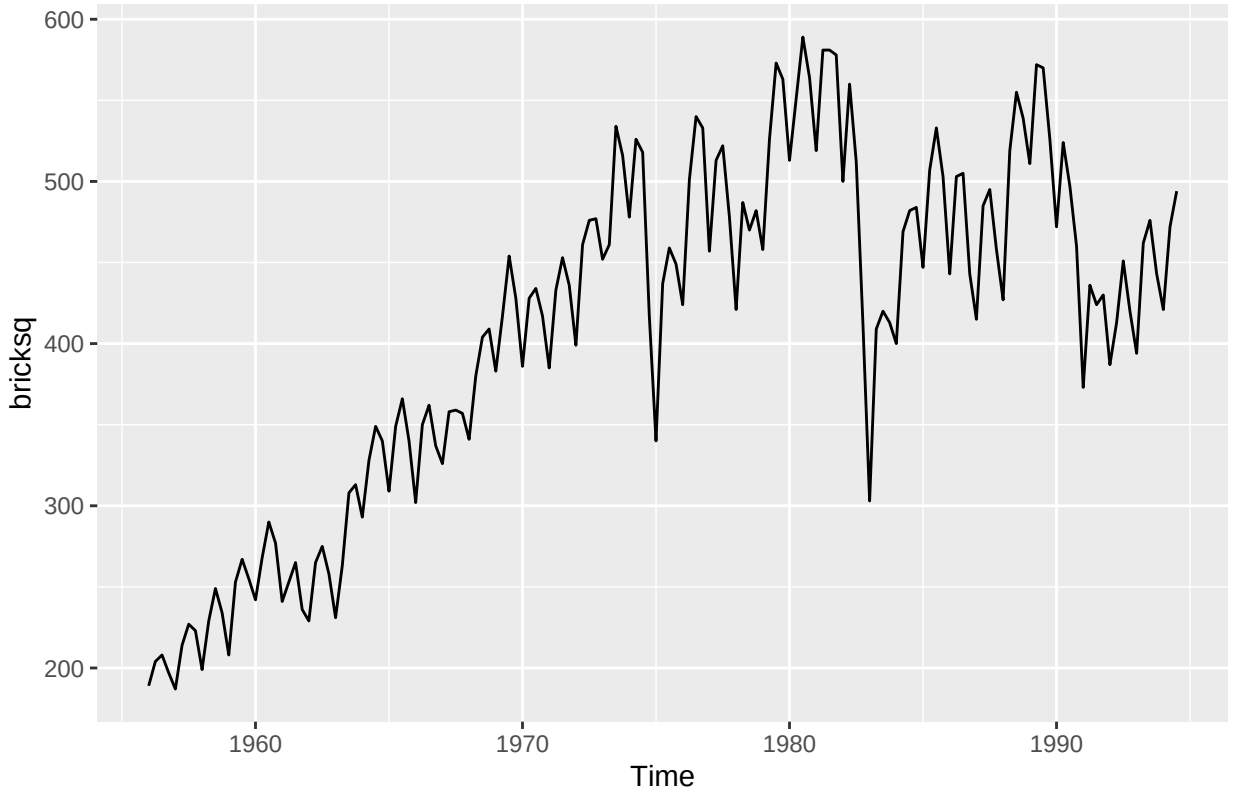
US has highest accidental deaths in July.

bricksq

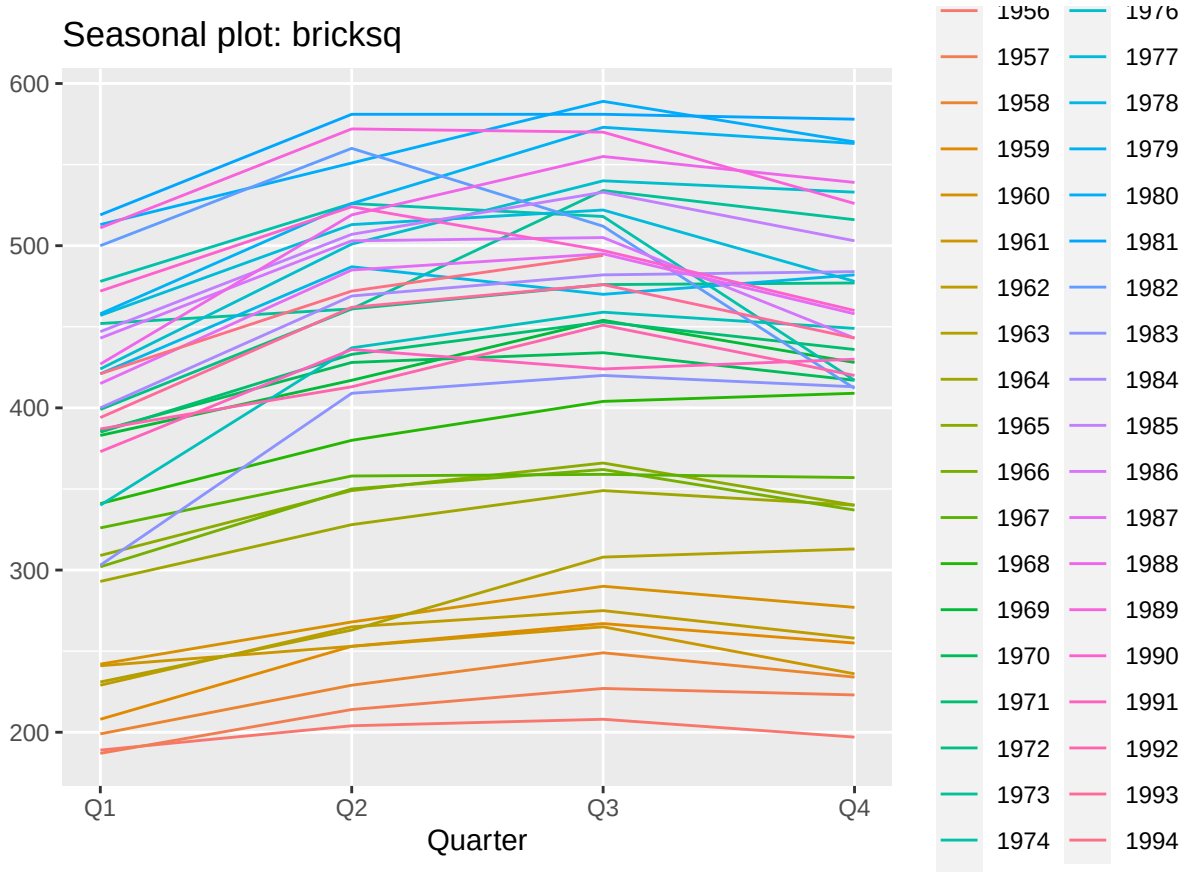
bricksq - Australian quarterly clay brick production: 1956–1994.

```
?bricksq
```

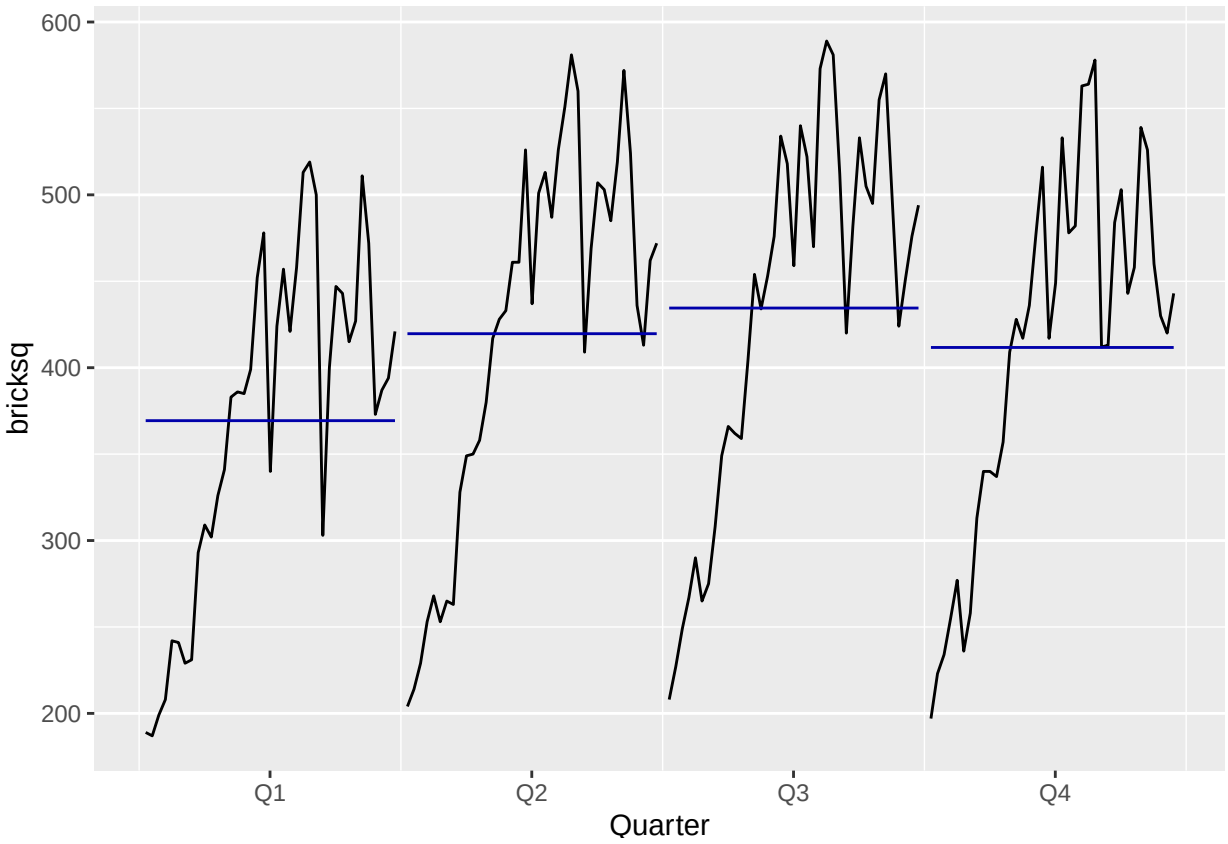
```
autoplot(bricksq)
```



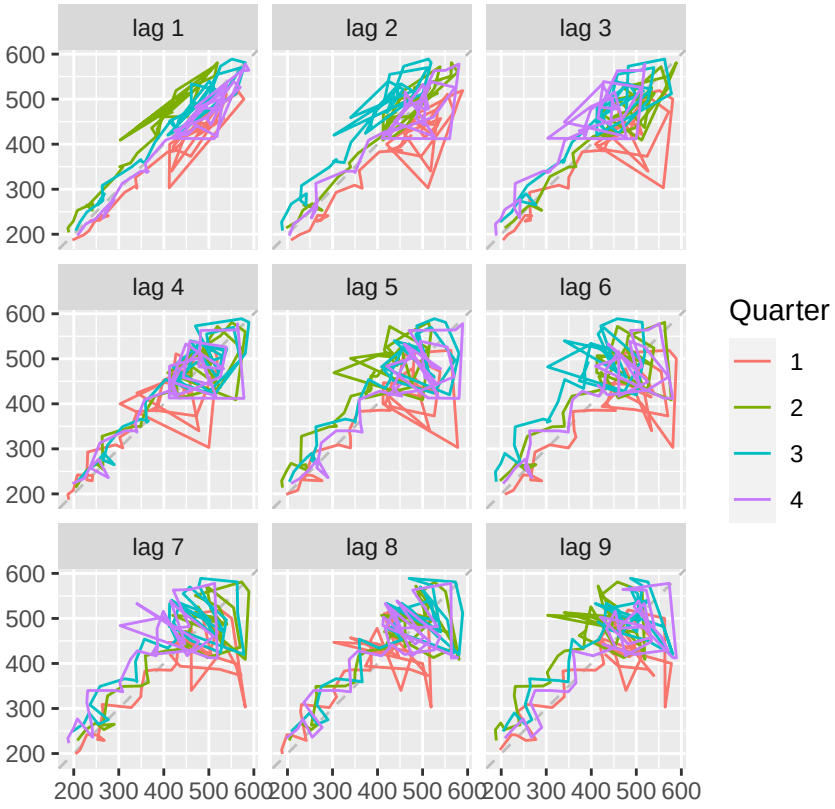

```
ggseasonplot(bricksq)
```



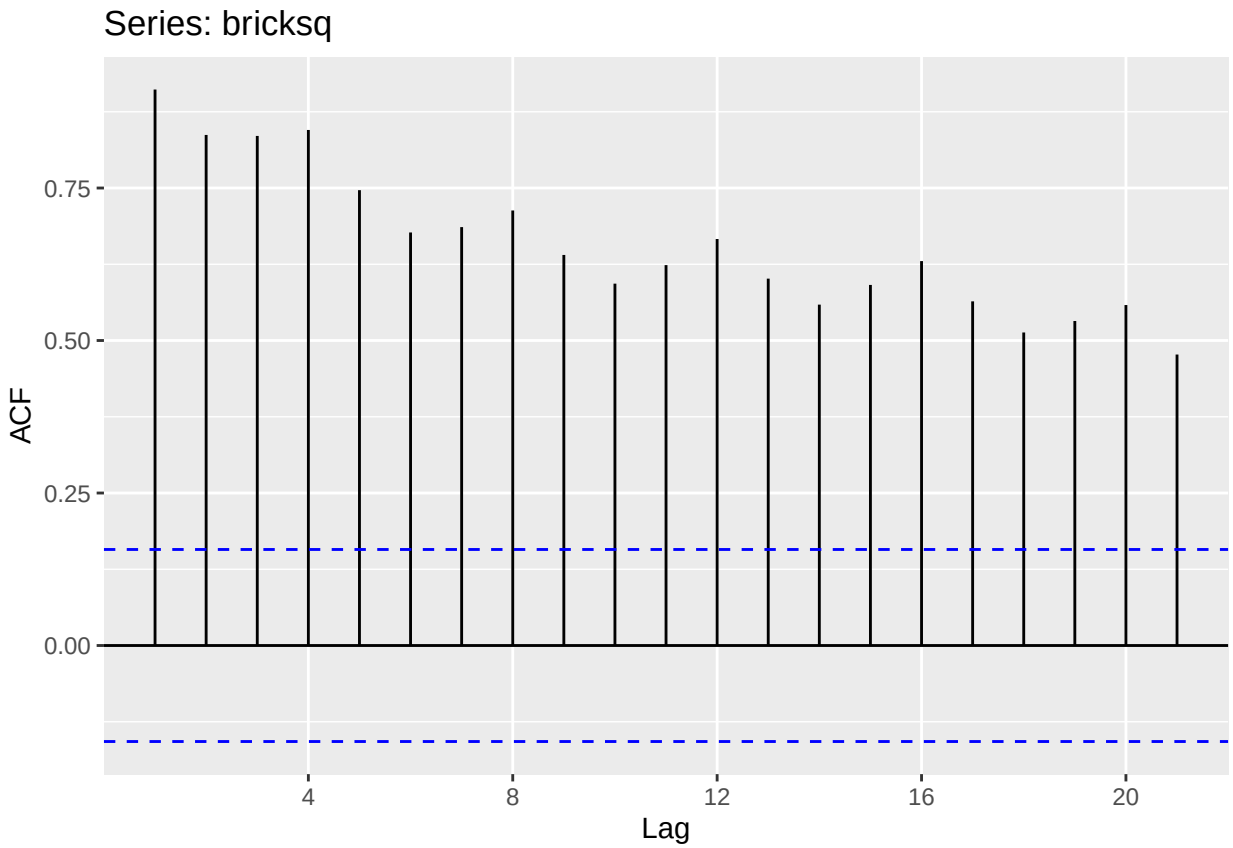

```
ggsubseriesplot(bricksq)
```




```
gglagplot(bricksq)
```



```
ggAcf(bricksq)
```



a. Can you spot any seasonality, cyclicity and trend?

Annual seasonality with peak at Q3 and trough at Q1. Generally it shows an upward trend till 1980. I also see a cyclic behavior in 8 year. The lagplot shows string linear relationship at lag 1. The slow decrease in the ACF as the lags increase is due to the trend.

b. What do you learn about the series?

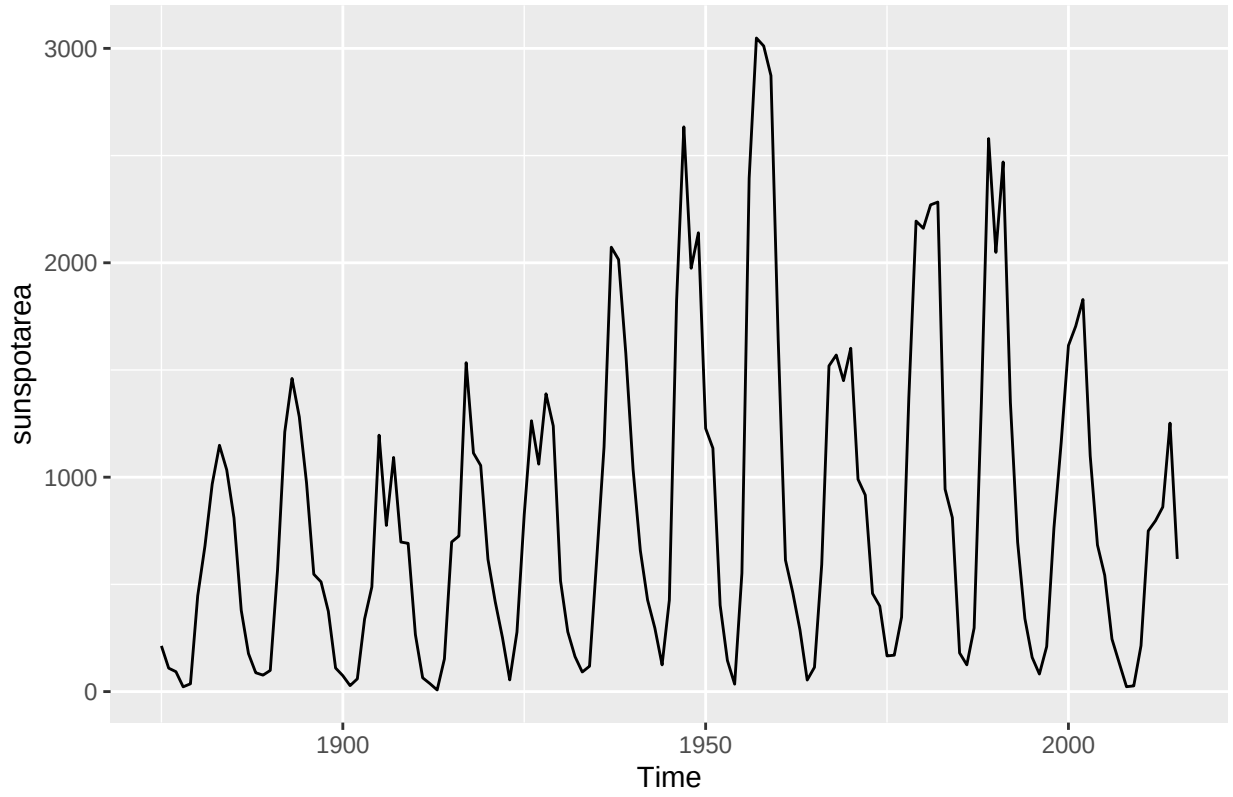
Australian quarterly clay brick production is lowest at Q1 and highest at Q3.

sunspotarea

sunspotarea - Annual averages of the daily sunspot areas (in units of millionths of a hemisphere) for the full sun.

```
?sunspotarea
```

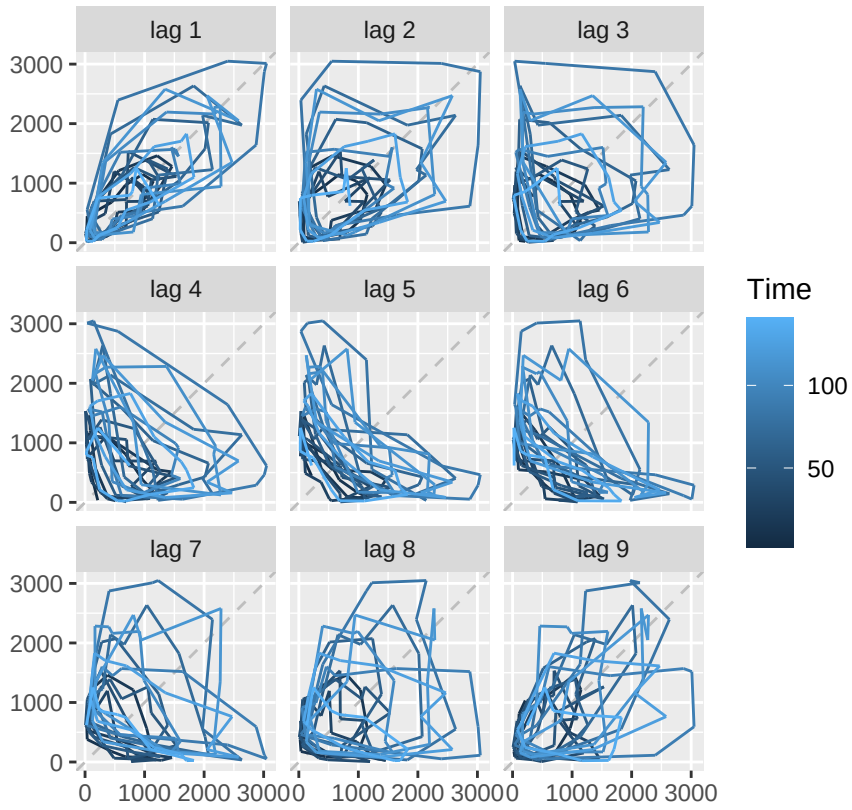
```
autoplot(sunspotarea)
```



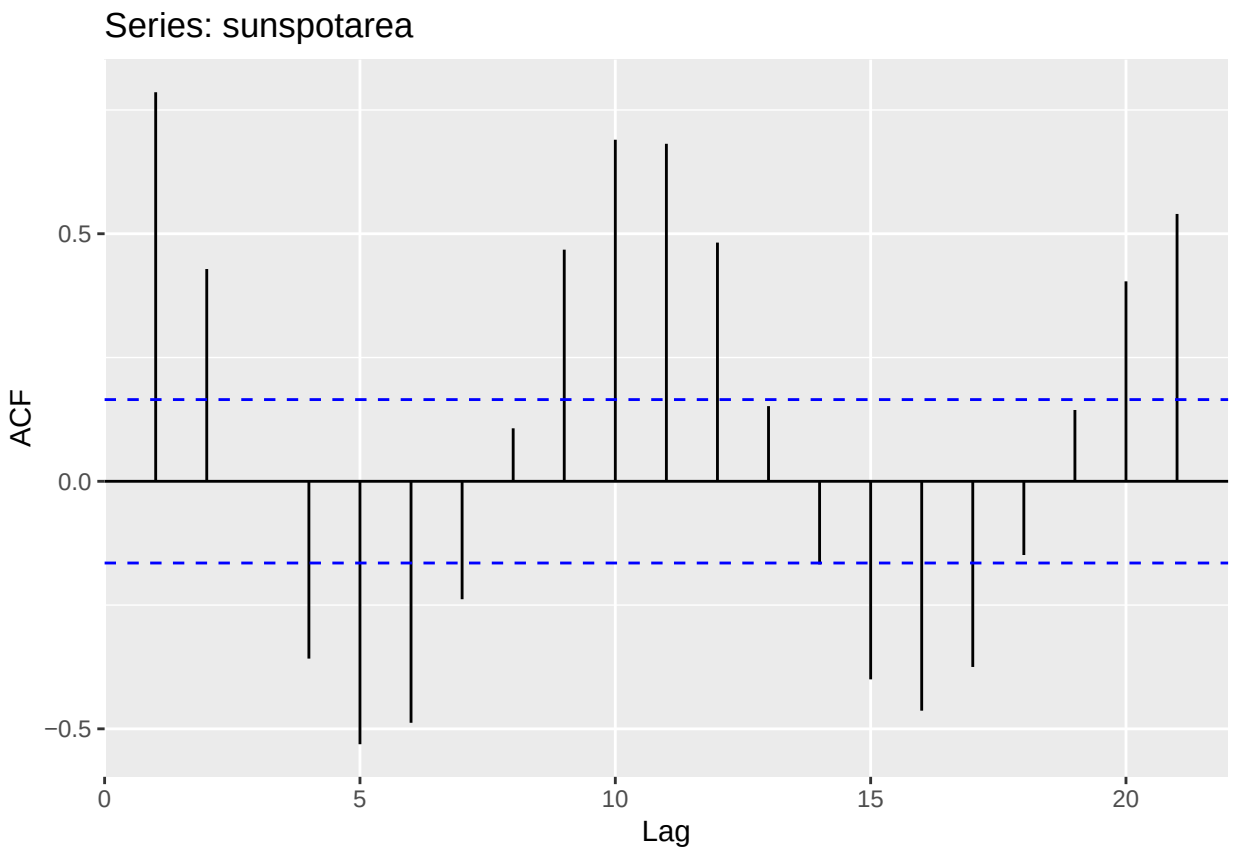

```
#ggseasonplot(sunspotarea)
# Data are not seasonal
```

```
#ggsubseriesplot(sunspotarea)
# Data are not seasonal
```

```
gglagplot(sunspotarea)
```




```
ggAcf(sunspotarea)
```



a. Can you spot any seasonality, cyclicity and trend?

Don't see any trend and seasonality in plots. It appears to have cyclicity in about 10-12 years. The acf plot has positive and negative correlation peaks which confirms that each cycle in 10 - 12 years.

b. What do you learn about the series?

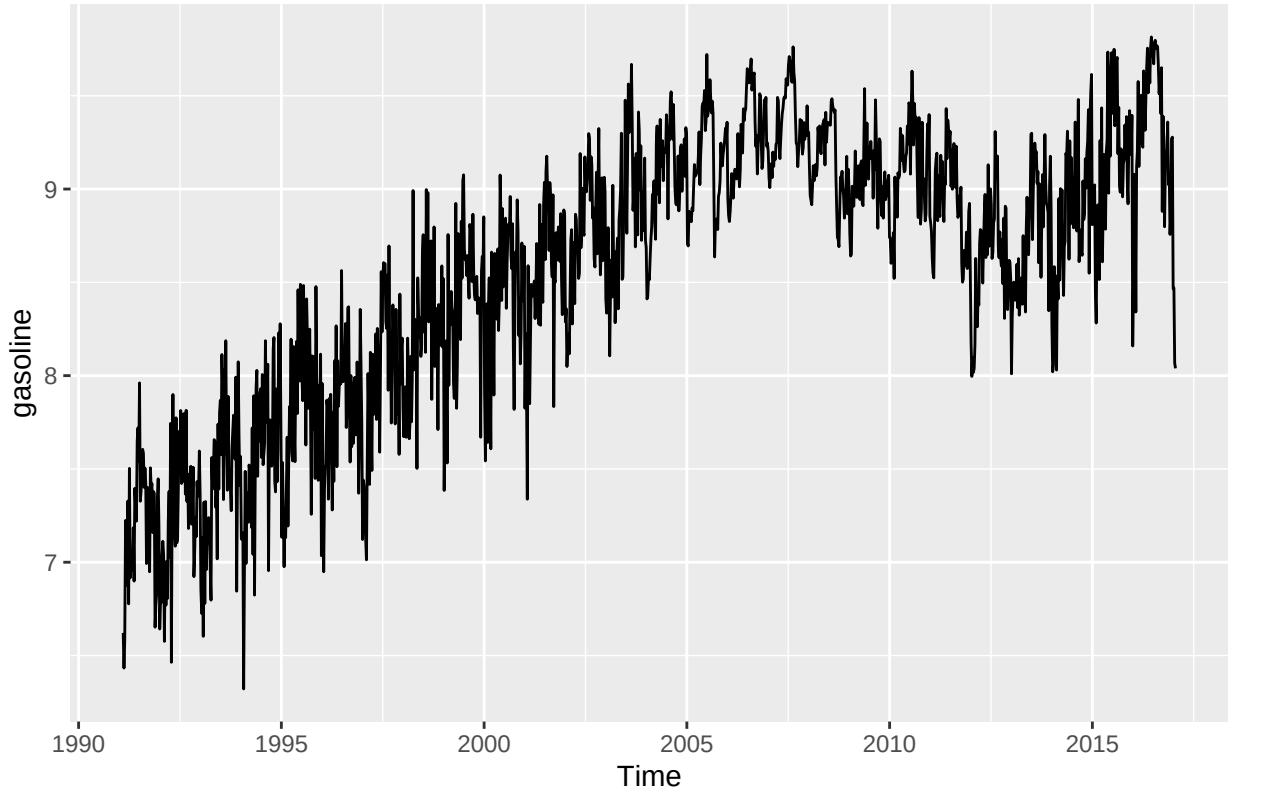
This time series does have cycle in about 10-12 years.

gasoline

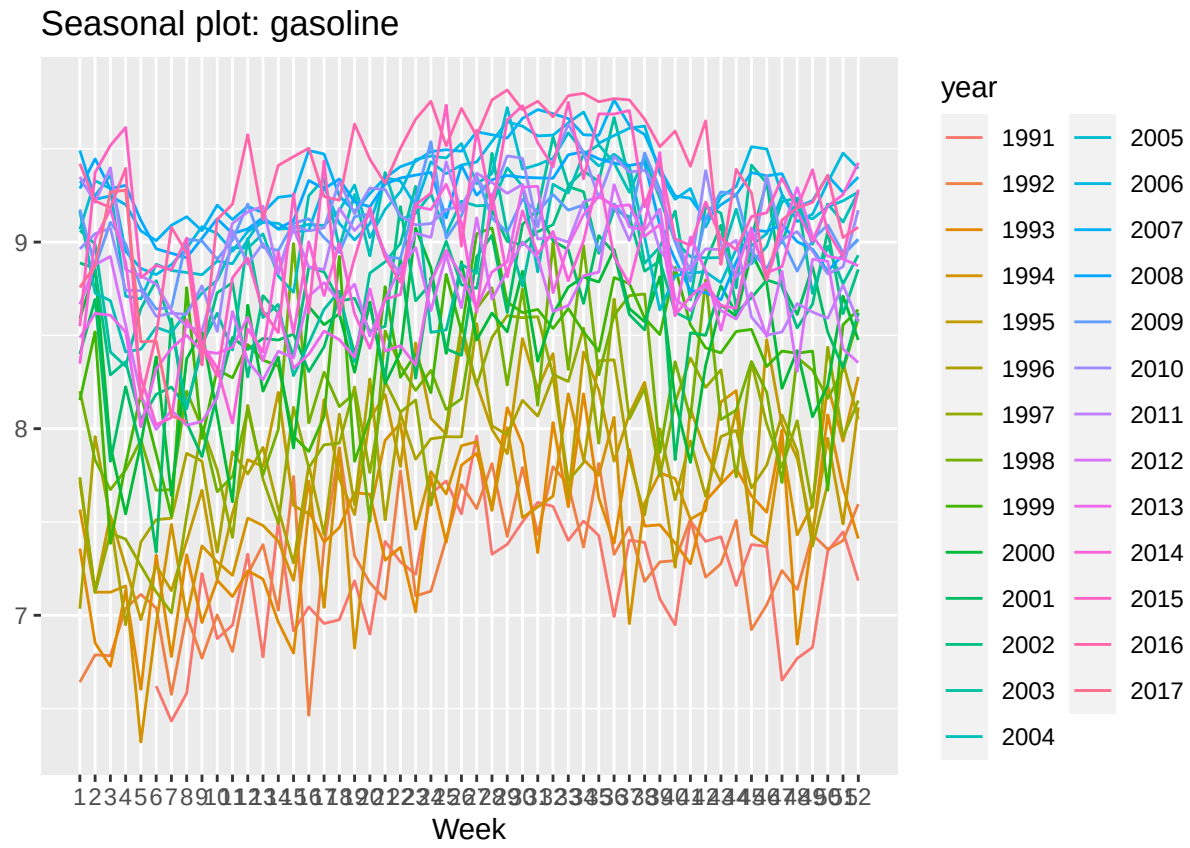
gasoline - Weekly data beginning 2 February 1991, ending 20 January 2017. Units are "million barrels per day".

```
?gasoline
```

```
autoplot(gasoline)
```

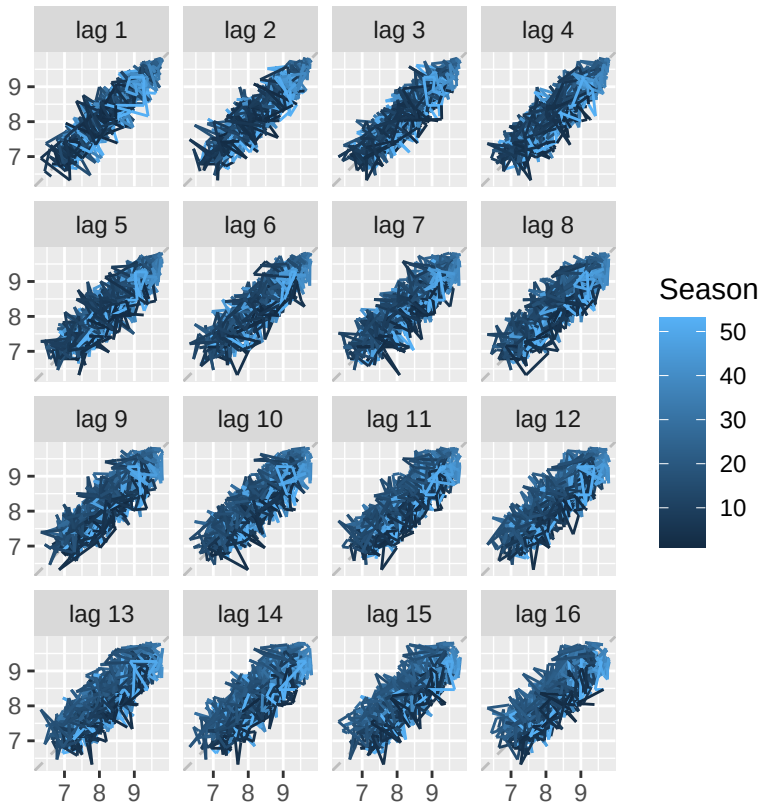



```
ggseasonplot(gasoline)
```

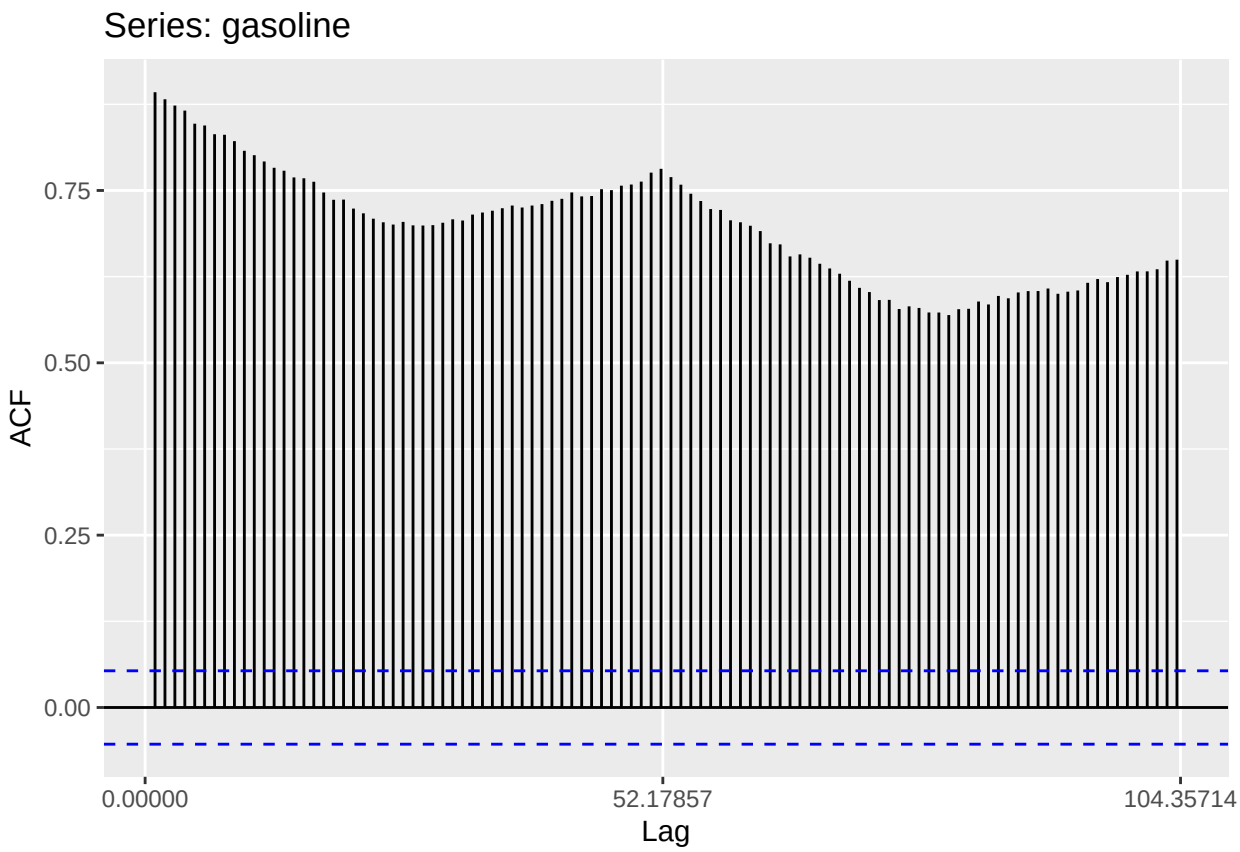



```
#ggsubseriesplot(gasoline)
# Each season requires at least 2 observations. This may be caused from specifying a time-series with n
```

```
gglagplot(gasoline)
```




```
ggAcf(gasoline)
```



a. Can you spot any seasonality, cyclicity and trend?

This time series shows upwards trend in general and has annual seasonality. Lagplot shows strongest linear relationship at lag 1.

b. What do you learn about the series?

In general, the trend of gasoline supply has been going up.