DATA624 - 2.10

Calvin Wong

2/5/2020

```r
library(fpp2)
```

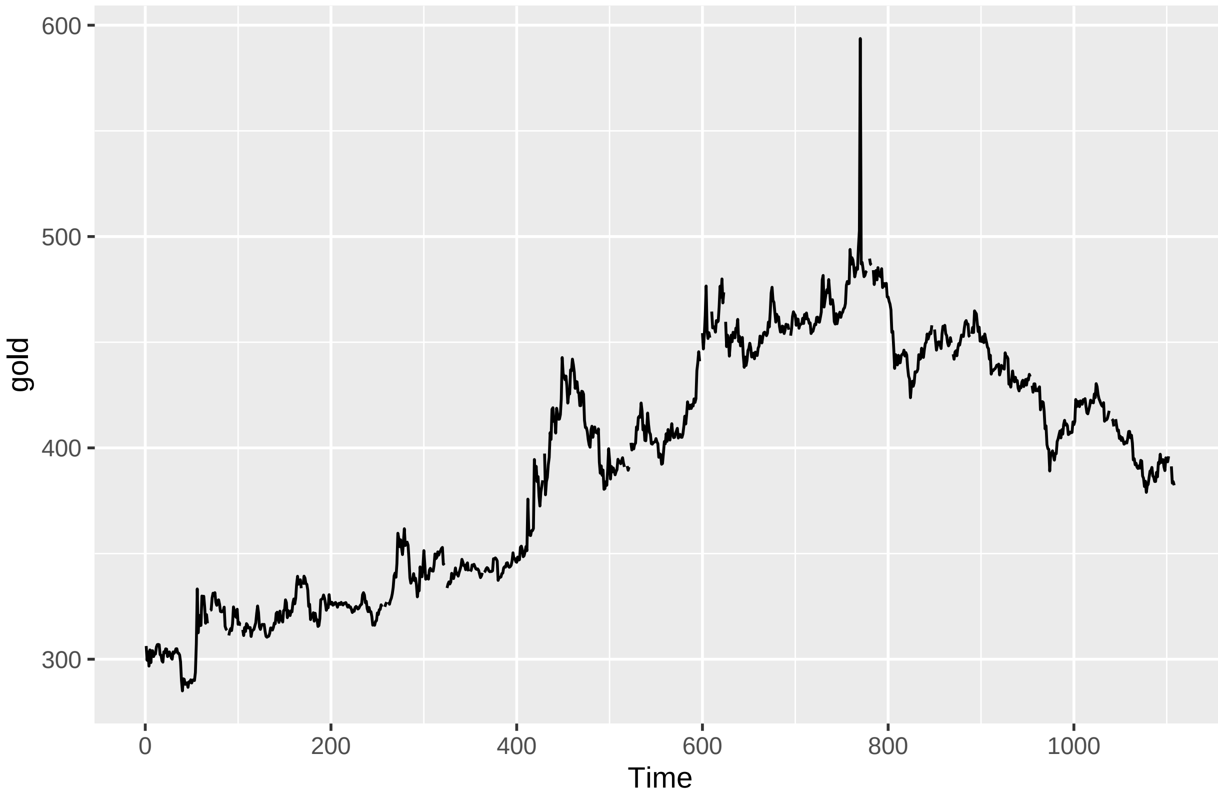
```
## Loading required package: ggplot2

## Warning: package 'ggplot2' was built under R version 3.5.2

## Loading required package: forecast

## Warning: package 'forecast' was built under R version 3.5.2

## Loading required package: fma

## Warning: package 'fma' was built under R version 3.5.2

## Loading required package: expsmooth
```

```r
library(forecast)
```

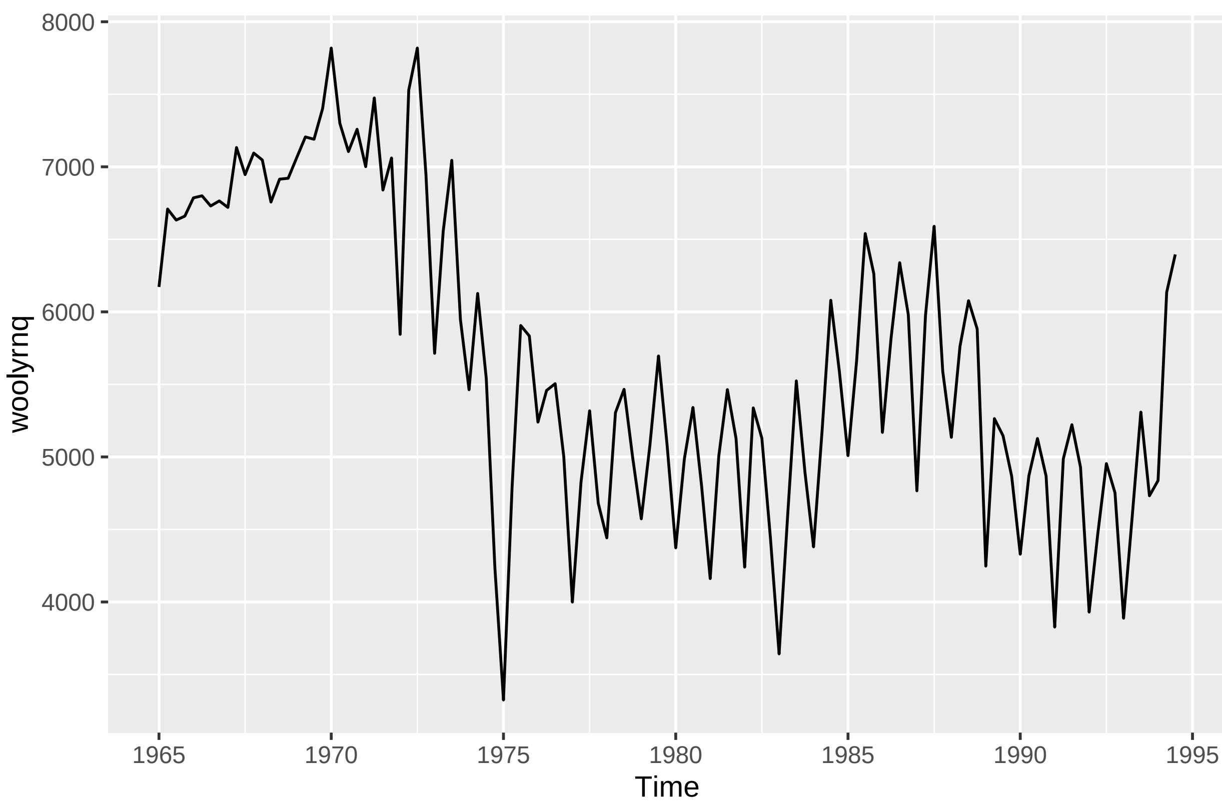
1) Use the help function to explore what the series gold, woolyrnq and gas represent.

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

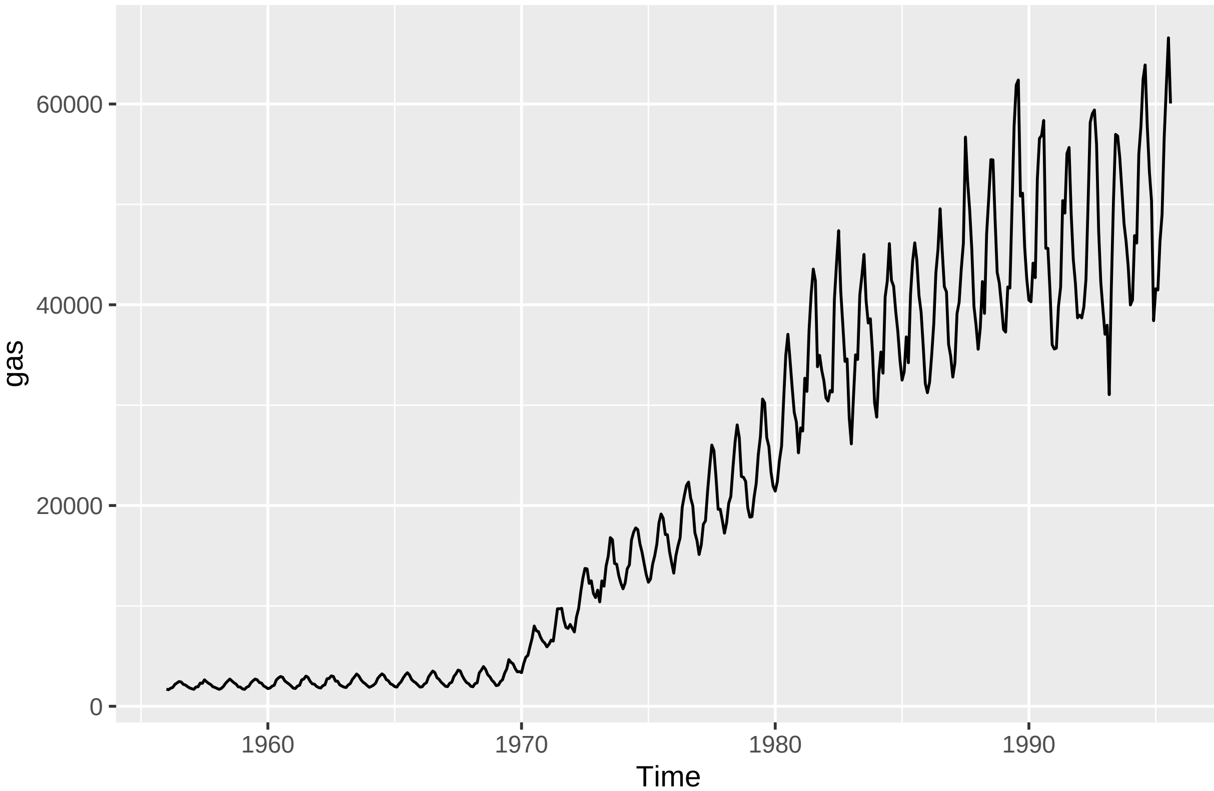
```r
autoplot(gold)
```




```r
autoplot(woolyrnq)
```



```r
autoplot(gas)
```



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

```r
frequency(gold)
```

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

```r
frequency(woolyrnq)
```

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

```r
frequency(gas)
```

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

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

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

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

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:

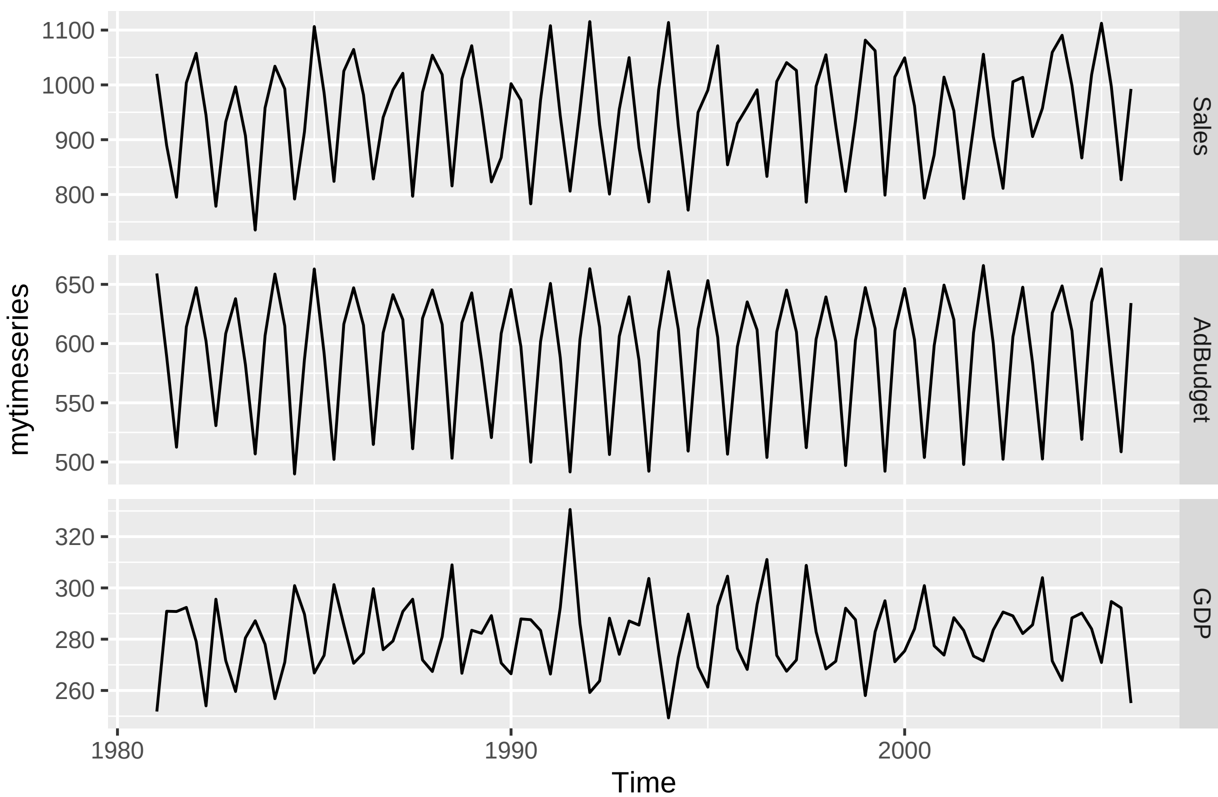
```
tute1 <- read.csv("https://otexts.com/fpp2/extrafiles/tute1.csv", header=TRUE)
```

b) Convert the data to time series

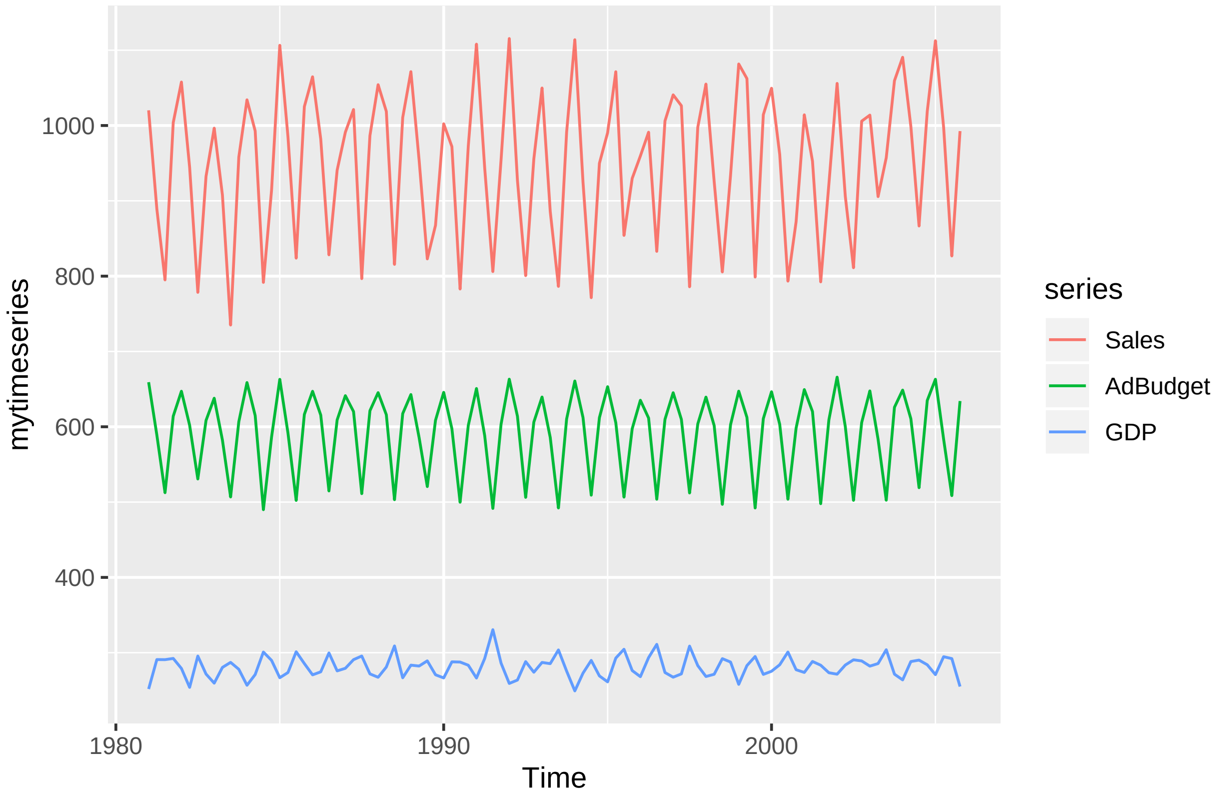
```
mytimeseries <- ts(tute1[,-1], start=1981, frequency=4)

##(The [,-1] removes the first column which contains the quarters as we don't need them now.)
```

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.
autoplot(mytimeseries)
```



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)
##The second argument (skip=1) is required because the Excel sheet has two header rows.
```

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

```
myts <- ts(retaildata[,"A3349909T"],
  frequency=12, start=c(1982,4))
```

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

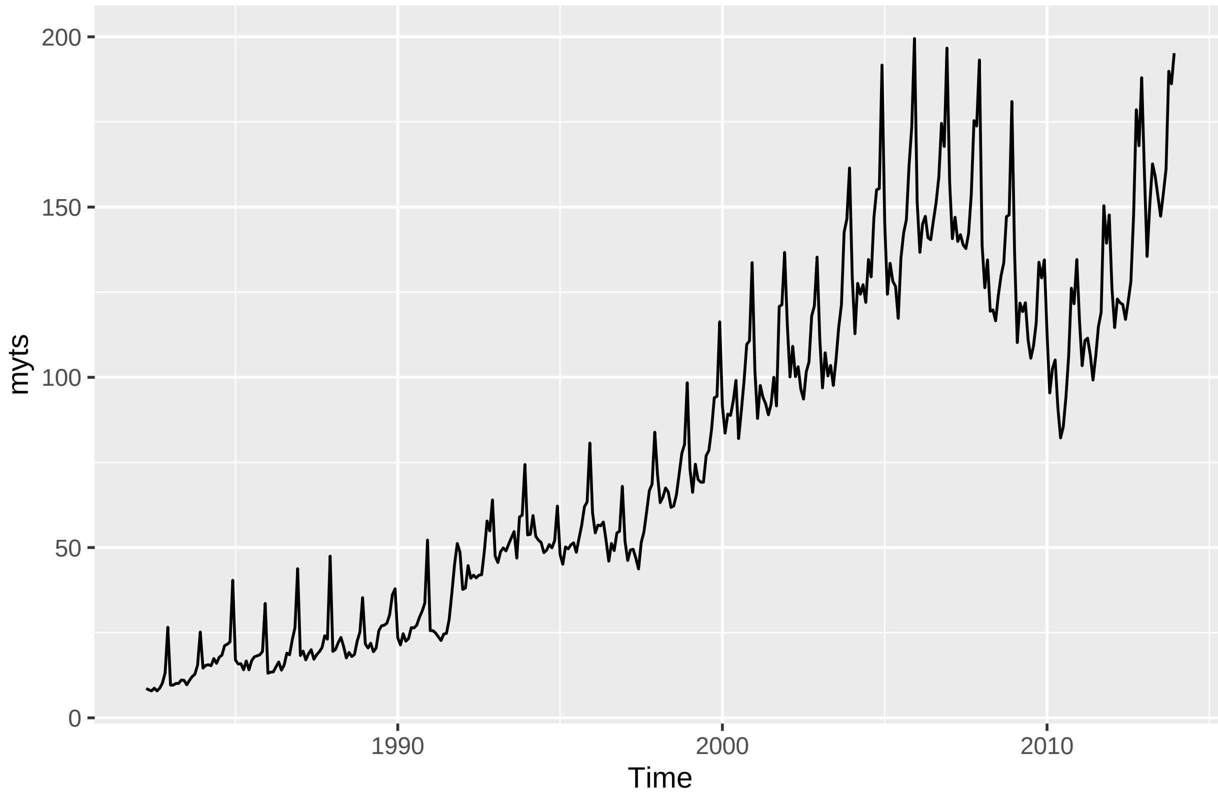
autoplot(), ggseasonplot(), ggsubseriesplot(), gglagplot(), ggAcf()

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

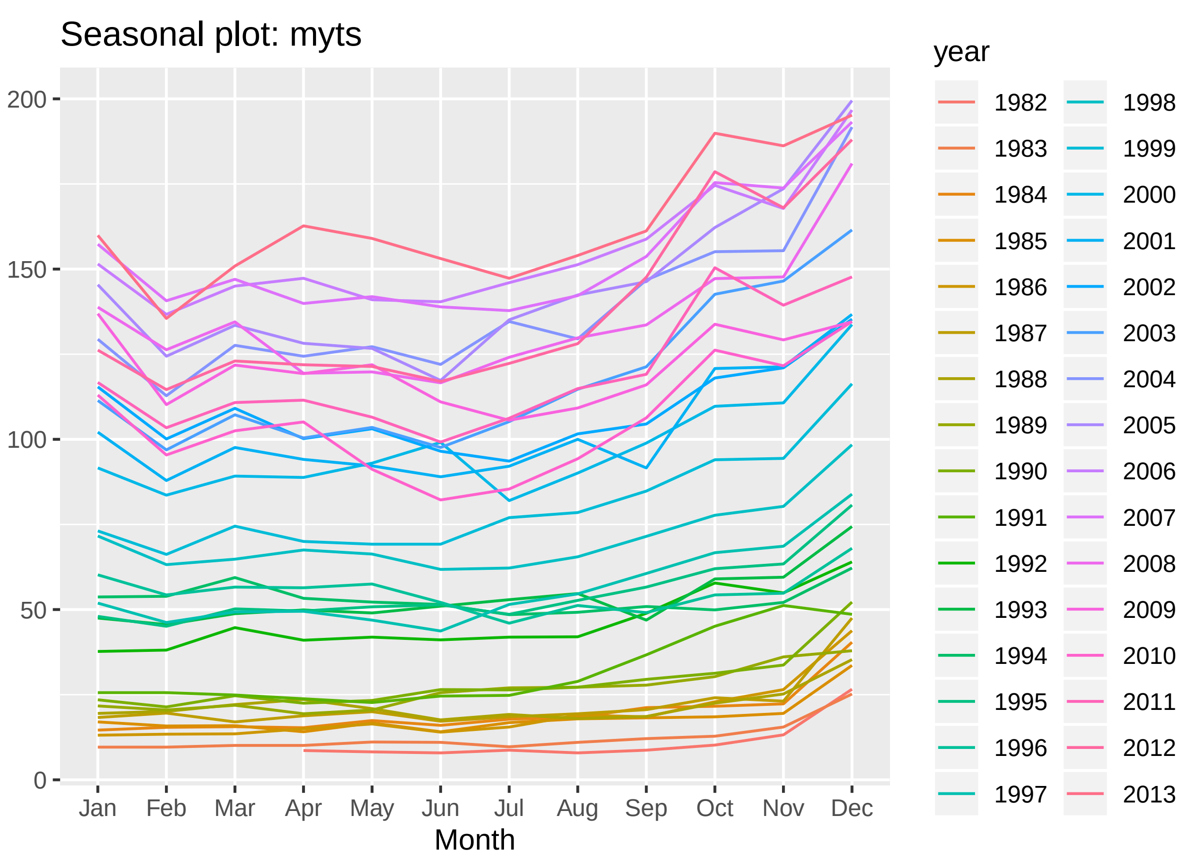
This data appears to have a positive trend due to the ACF trended time series that have slowly decreasing positive values over time. There does not seem to be a seasonal lag as evidenced in the ACF. Cyclical trend is not evident, maybe a larger dataset would show it.

Sales have been overall positively increasing (trend) over the timeline.

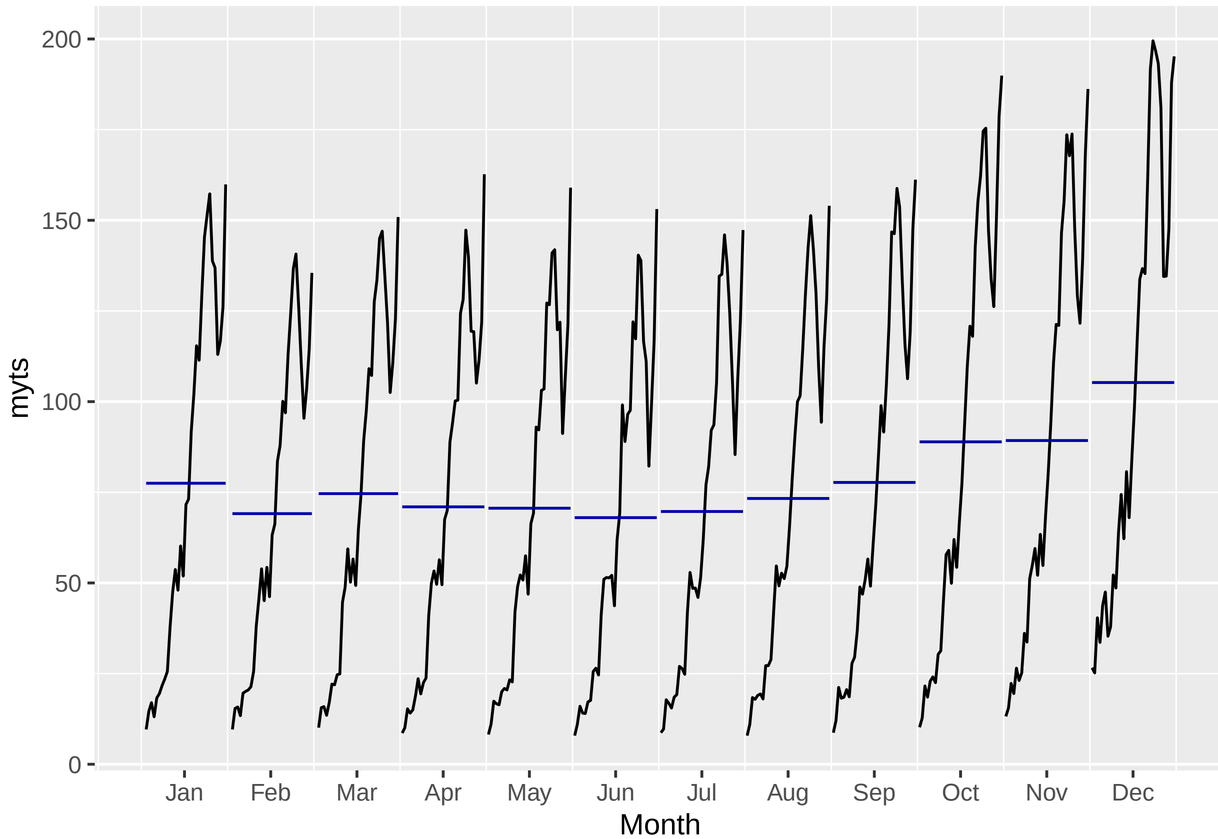
```
autoplot(myts)
```



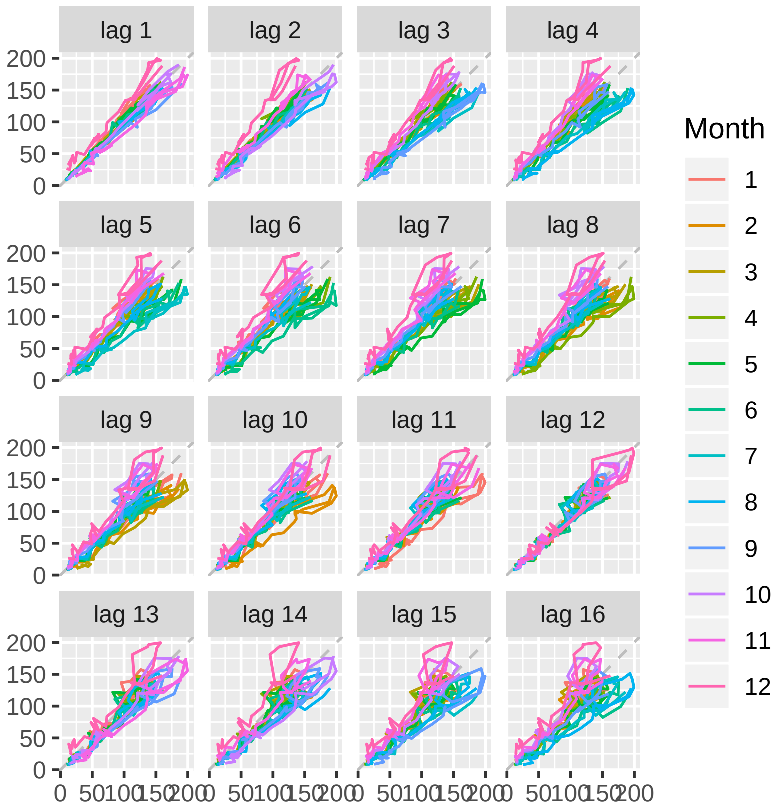

```
ggseasonplot(myts)
```



```r
ggsubseriesplot(myts)
```

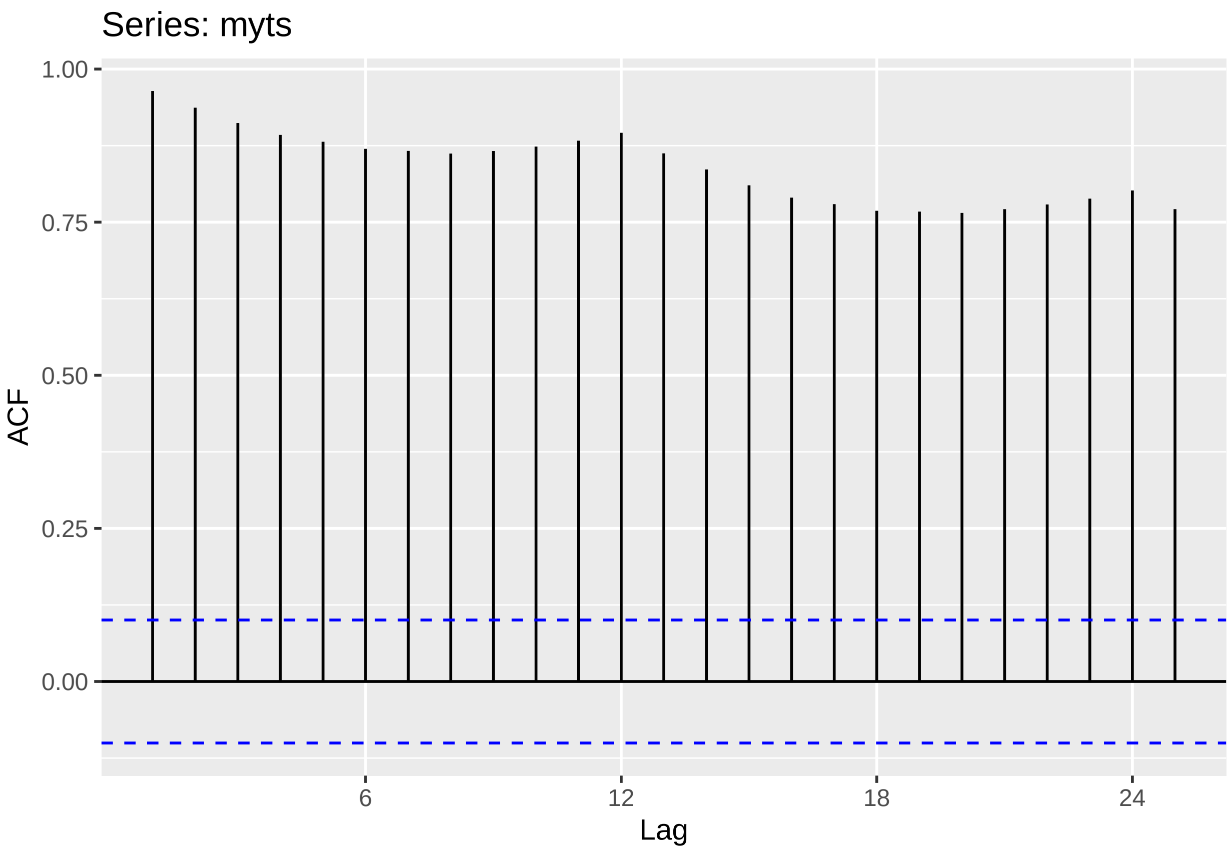



```r
gglagplot(myts)
```



lag 1
lag 2
lag 3
lag 4
lag 5
lag 6
lag 7
lag 8
lag 9
lag 10
lag 11
lag 12
lag 13
lag 14
lag 15
lag 16
0
50
100
150
200
Month
1
2
3
4
5
6
7
8
9
10
11
12


```r
ggAcf(myts)
```



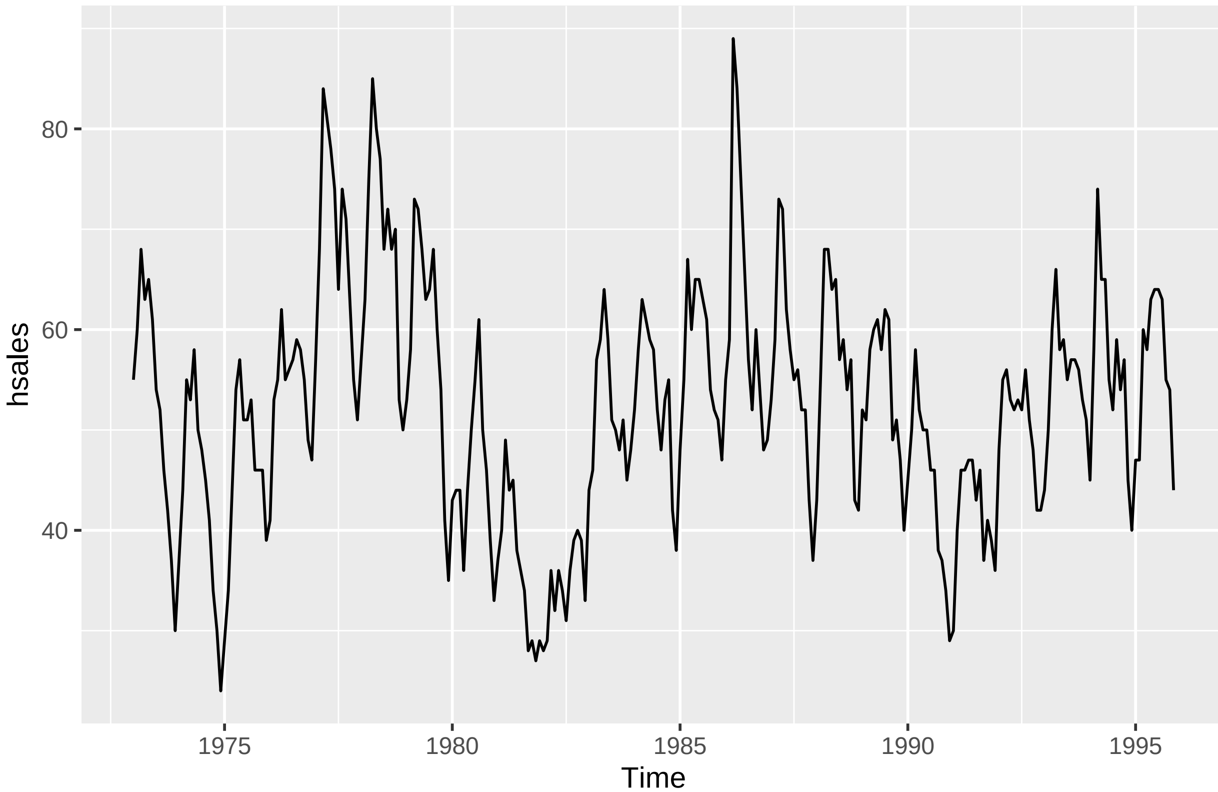
6) Use the following graphics functions: autoplot(), ggseasonplot(), ggsubseriesplot(), gglagplot(), ggAcf() and explore features from the following time series: hsales, usdeaths, bricksq, sunspotarea, gasoline.

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

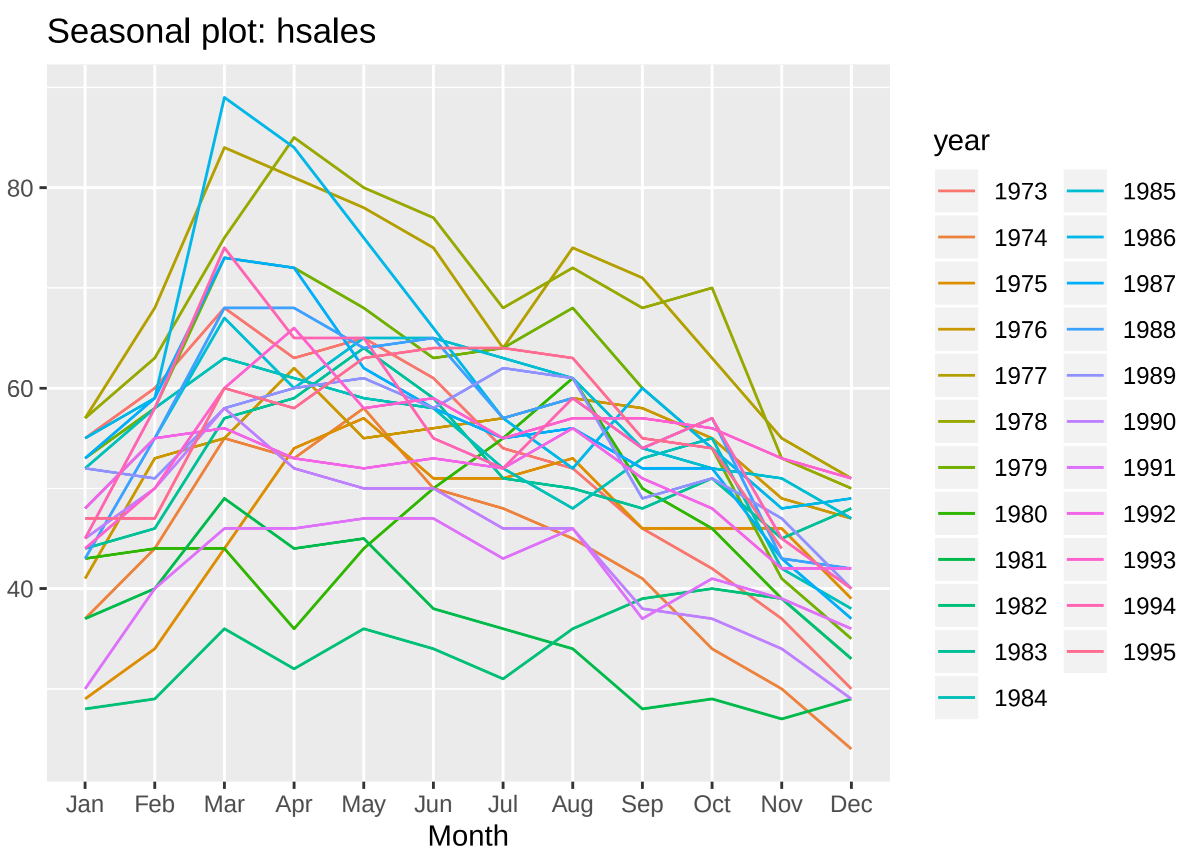
hsales

It seems that spring tends to sell the most amount of homes and there is a seasonal component to sales spikes. As the autocorrelation suggests, there appears to be a positive autocorrelation around 0, 12, 24. There may be a cyclical component as there appear to be some peaks and troughs cycling approximately every 12 years.

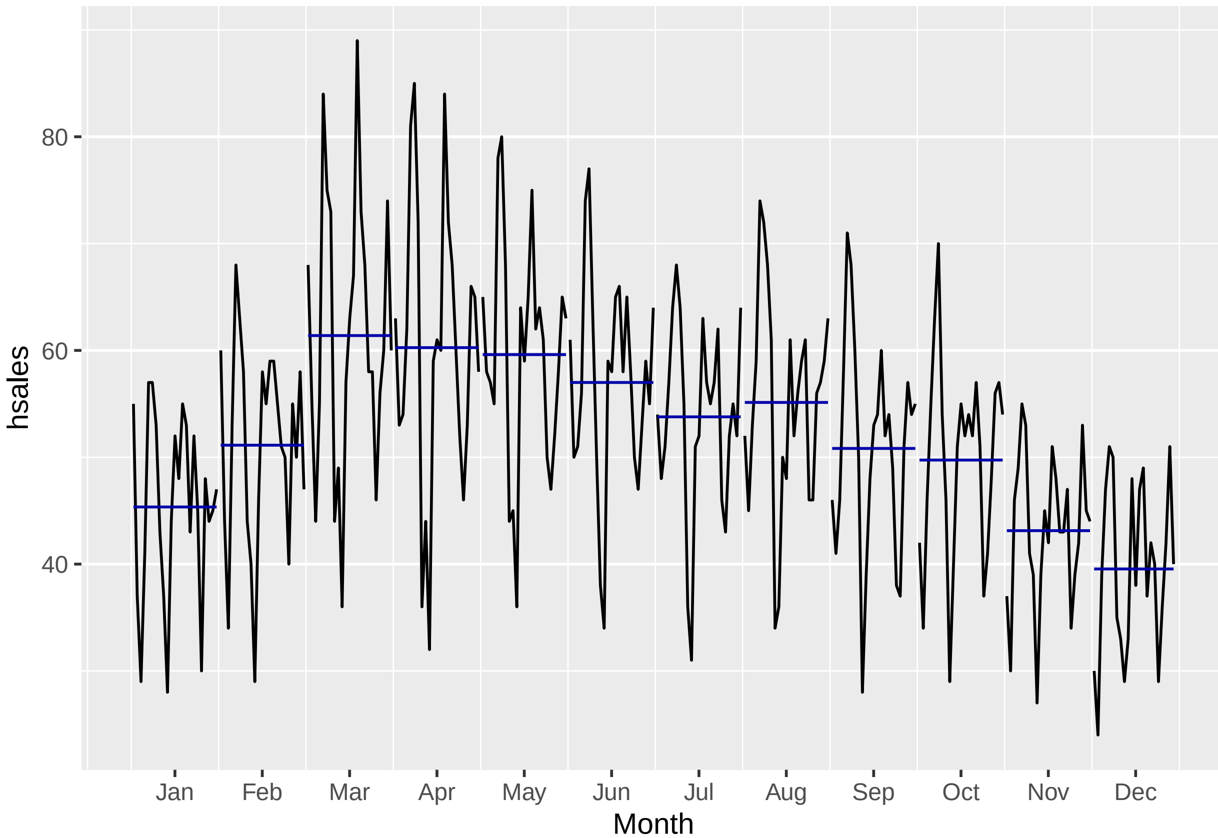
```
autoplot(hsales)
```



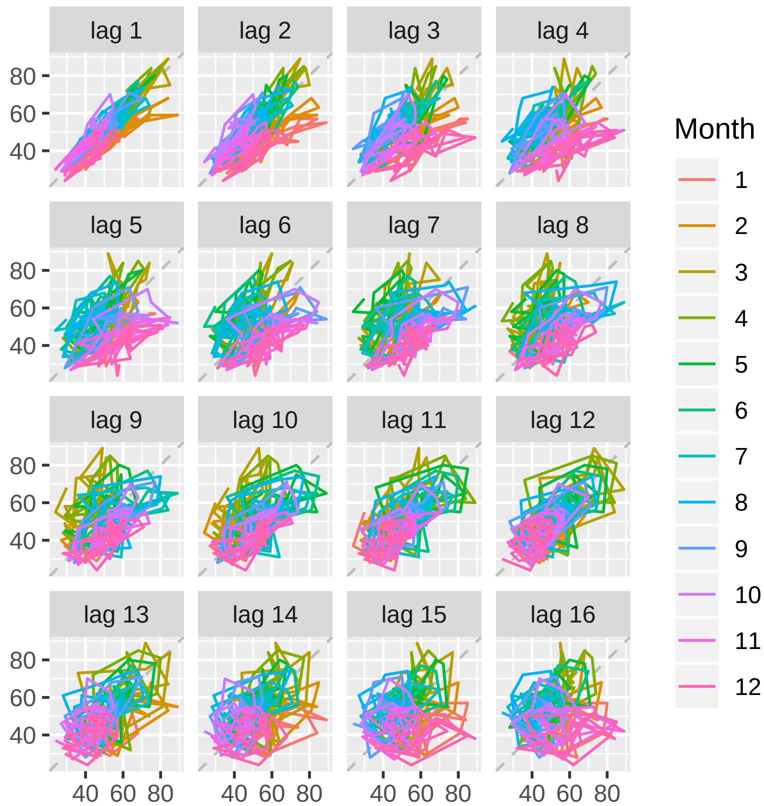

```
ggseasonplot(hsales)
```



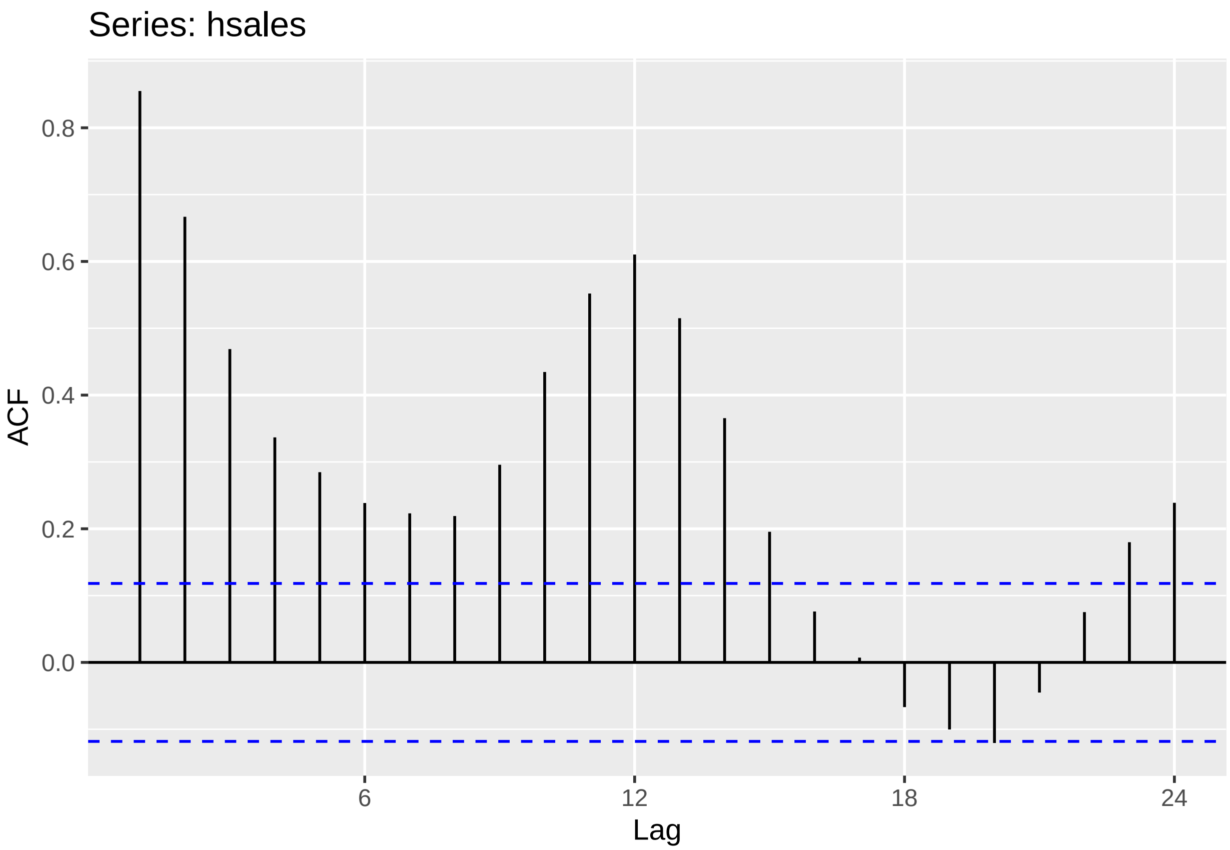
```r
ggsubseriesplot(hsales)
```




```r
gglagplot(hsales)
```



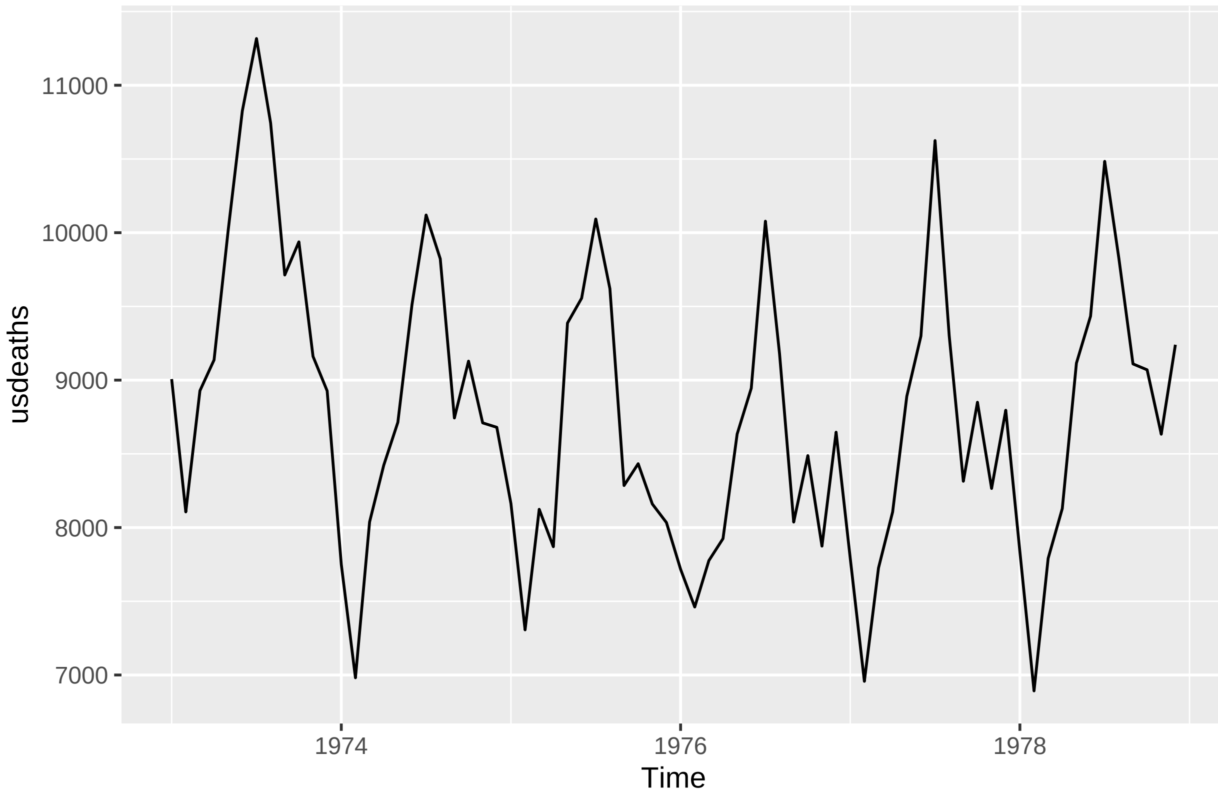

```
ggAcf(hsales)
```



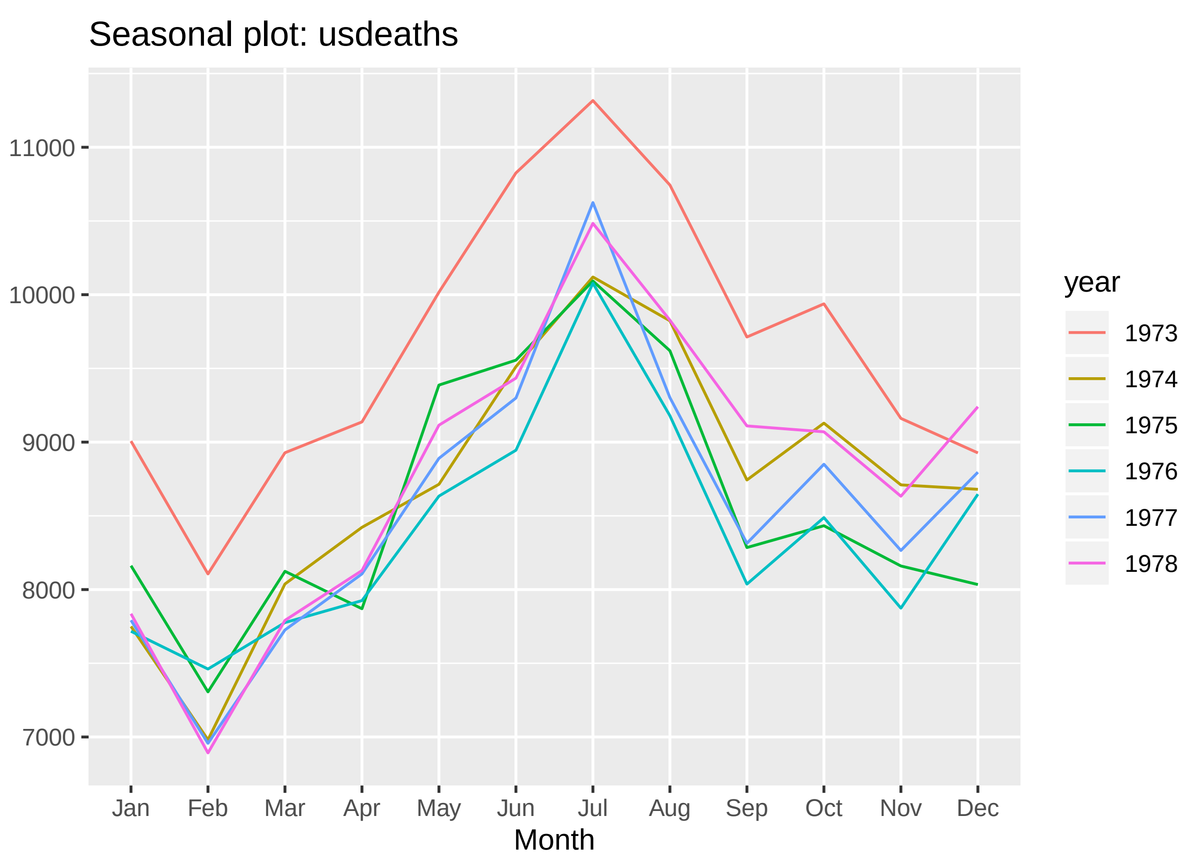
usdeaths

There certainly is a seasonal component (but no cyclical component). The number of accidental deaths occur spikes in the summer time. This correlates with the ACF and the lab plot. The lag plot shows the most positive correlation at lag 12 which corresponds to the 12th month.

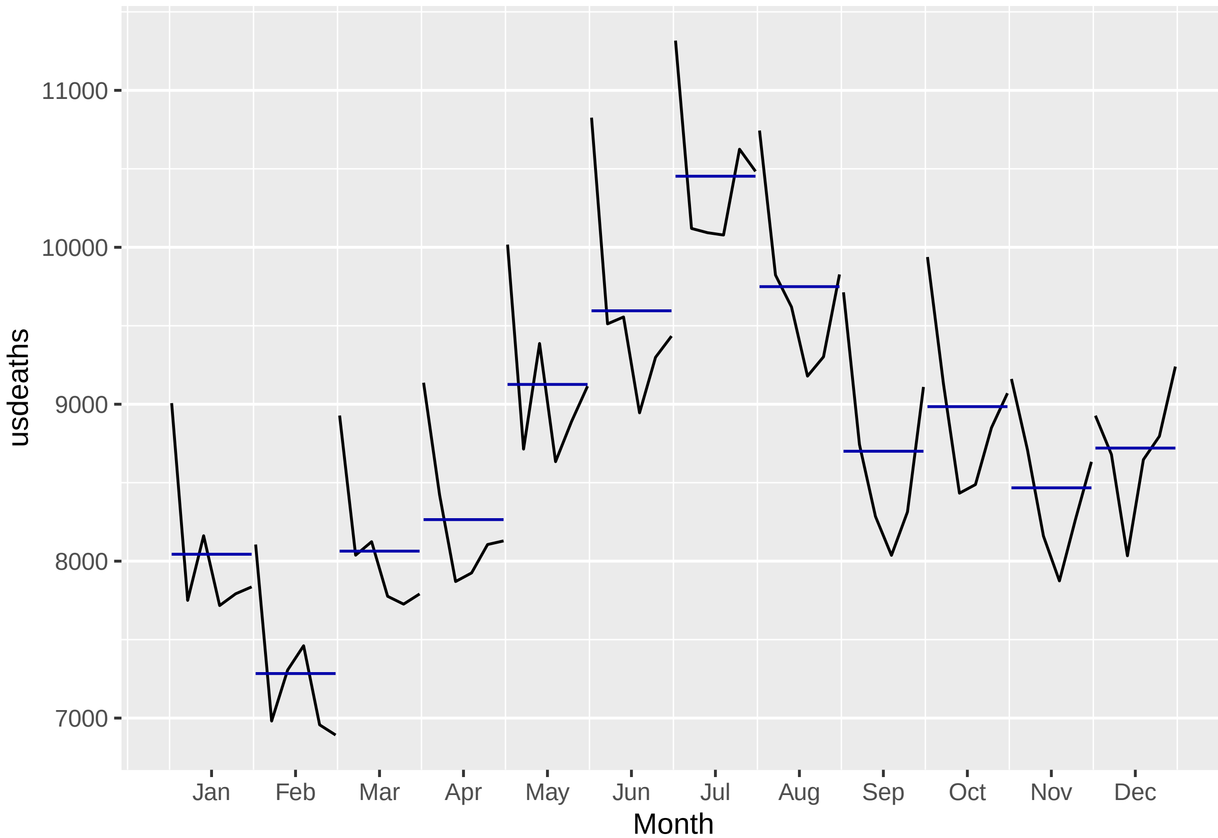
```r
autoplot(usdeaths)
```



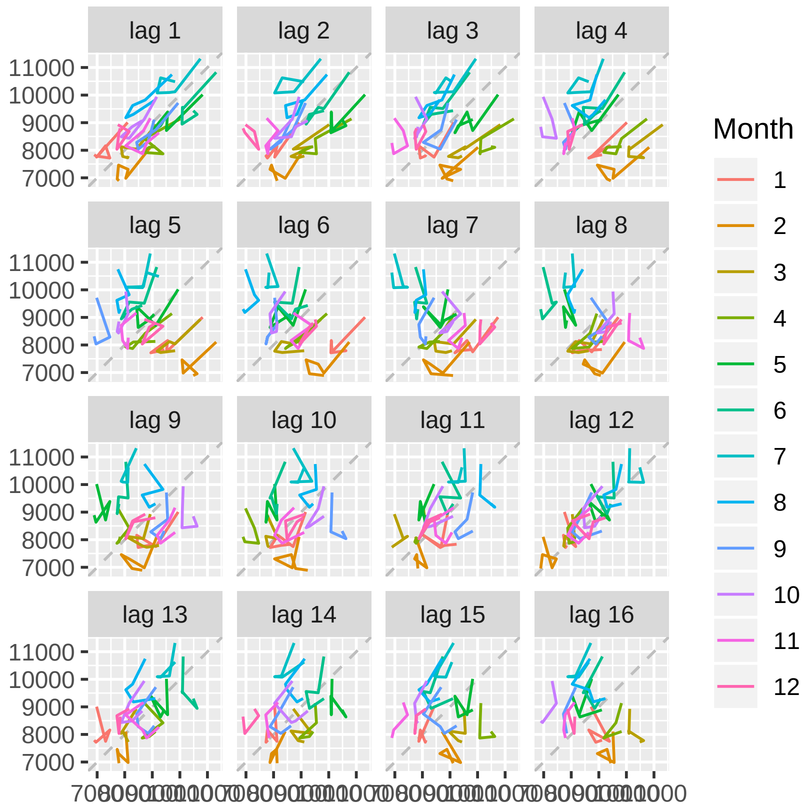

```
ggseasonplot(usdeaths)
```



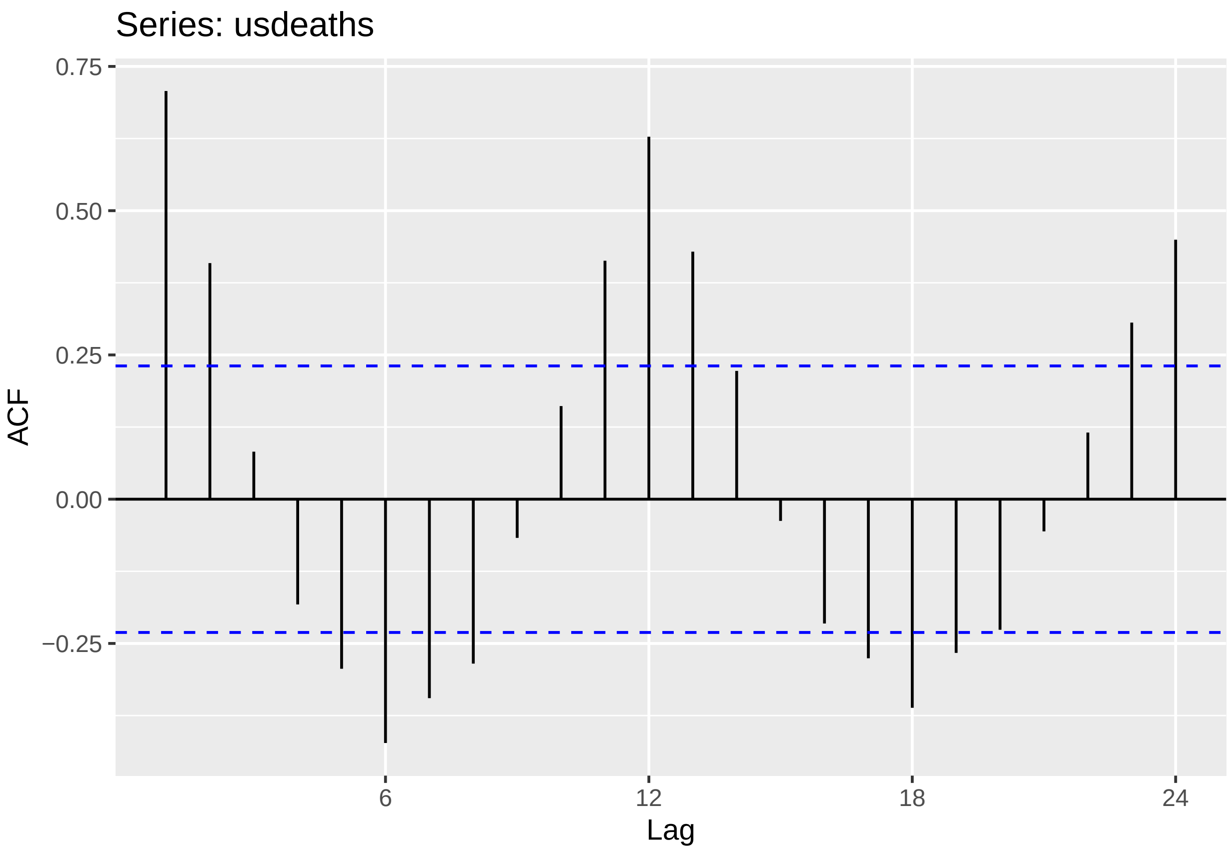
```r
ggsubseriesplot(usdeaths)
```




```r
gglagplot(usdeaths)
```



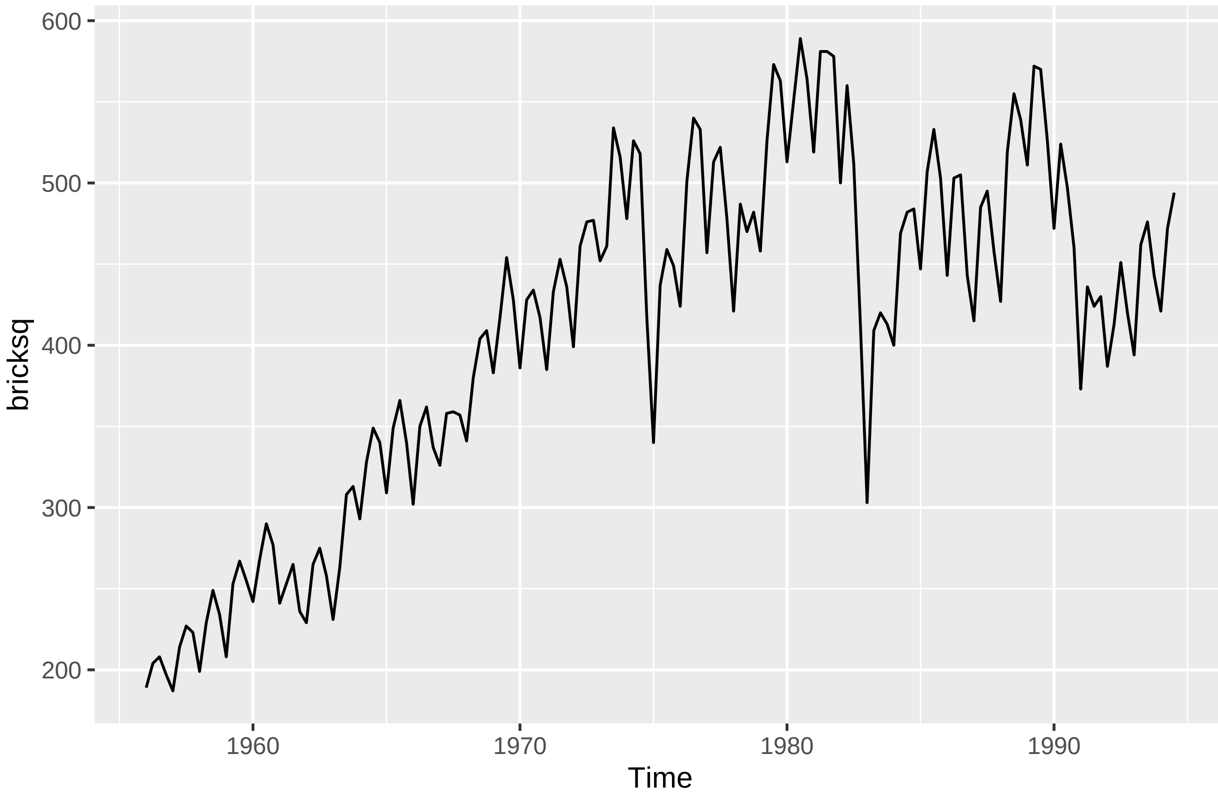

```r
ggAcf(usdeaths)
```



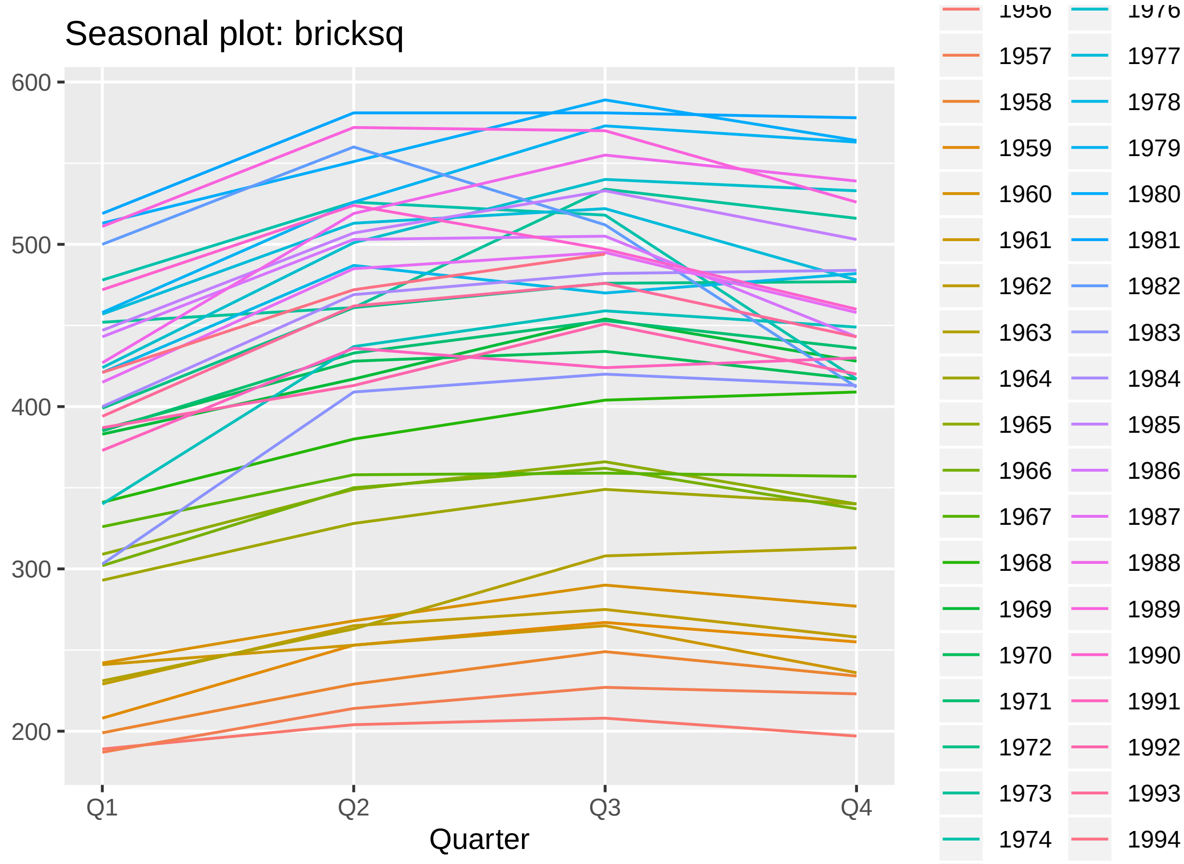
bricksq

There is a positive trend until year 1980 where it flattens. There are spikes at Q2 or Q3 yearly with a dip in Q4. This could coincide with construction activity peaking towards the summer months. The subseries plots also do not show significant difference between the quarters but there seems to be a pattern. The demand for brick clay seem to have stabilize towards the end of the century.

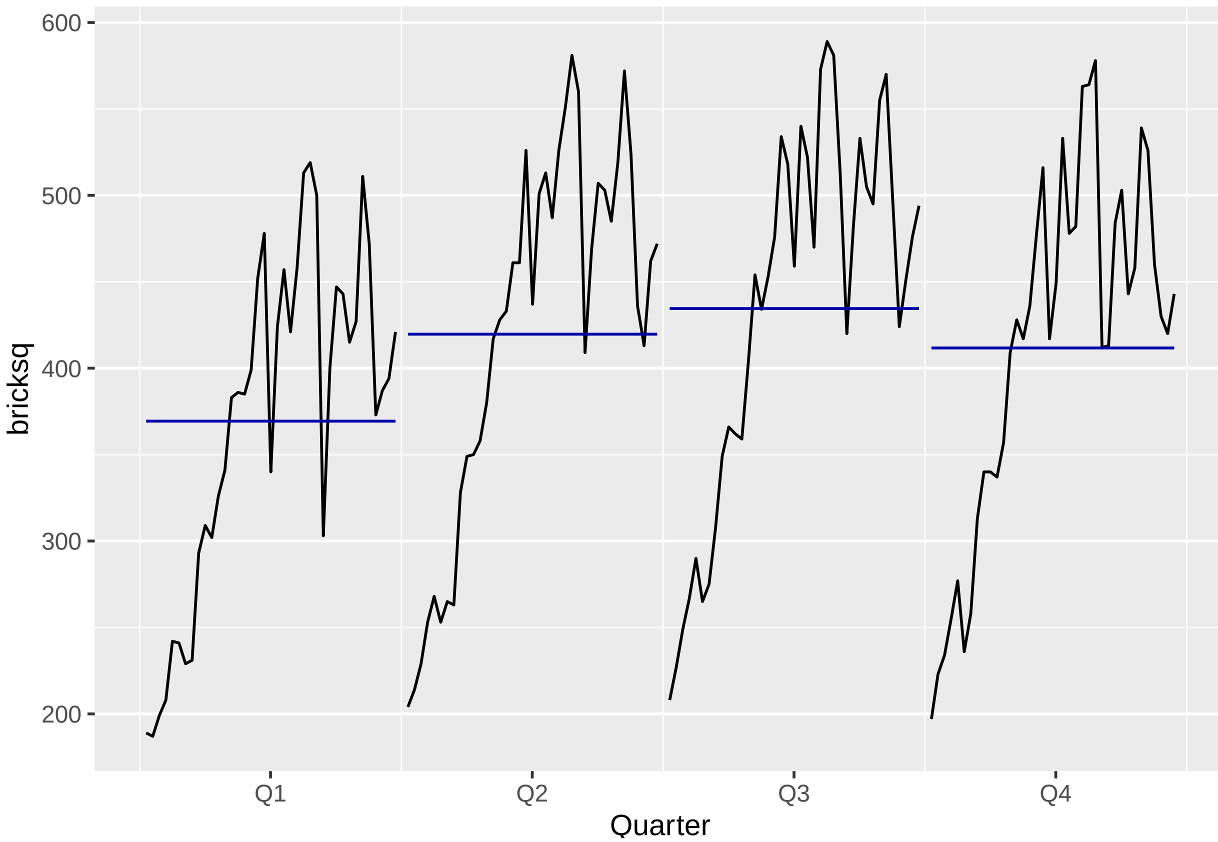
```r
autoplot(bricksq)
```



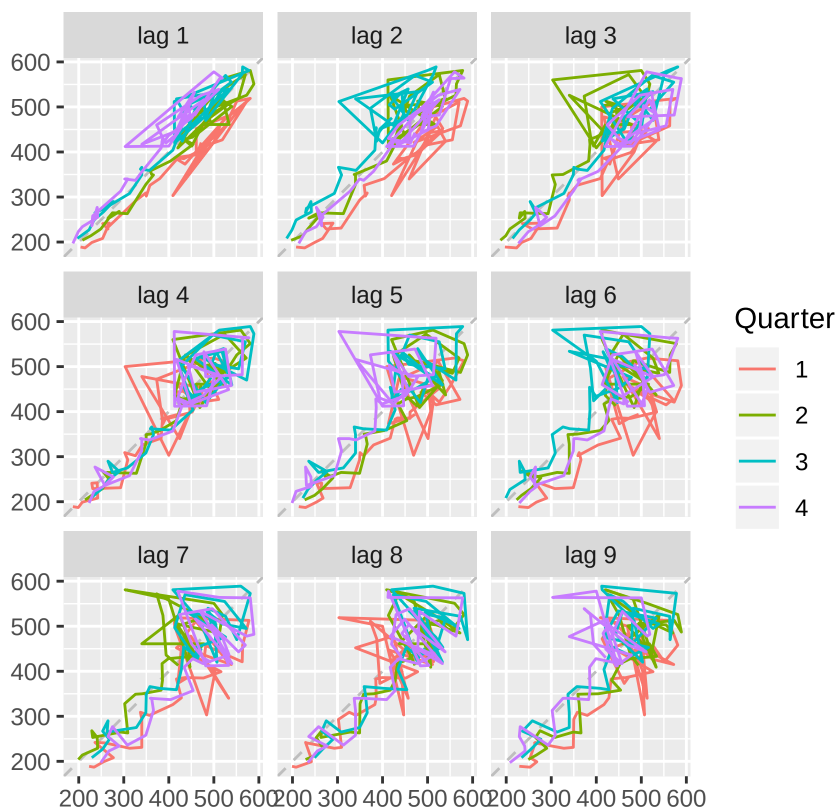

```r
ggseasonplot(bricksq)
```



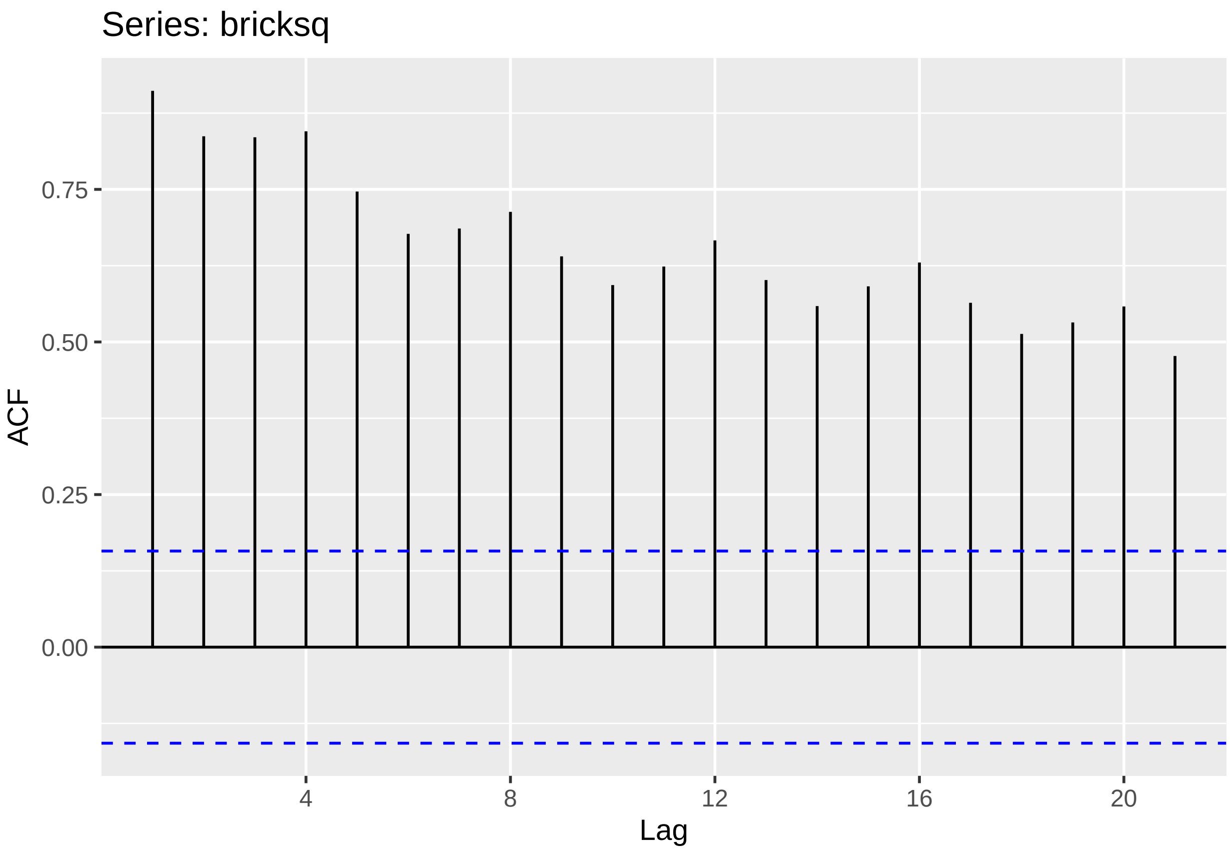

```
ggsubseriesplot(bricksq)
```




```r
gglagplot(bricksq)
```



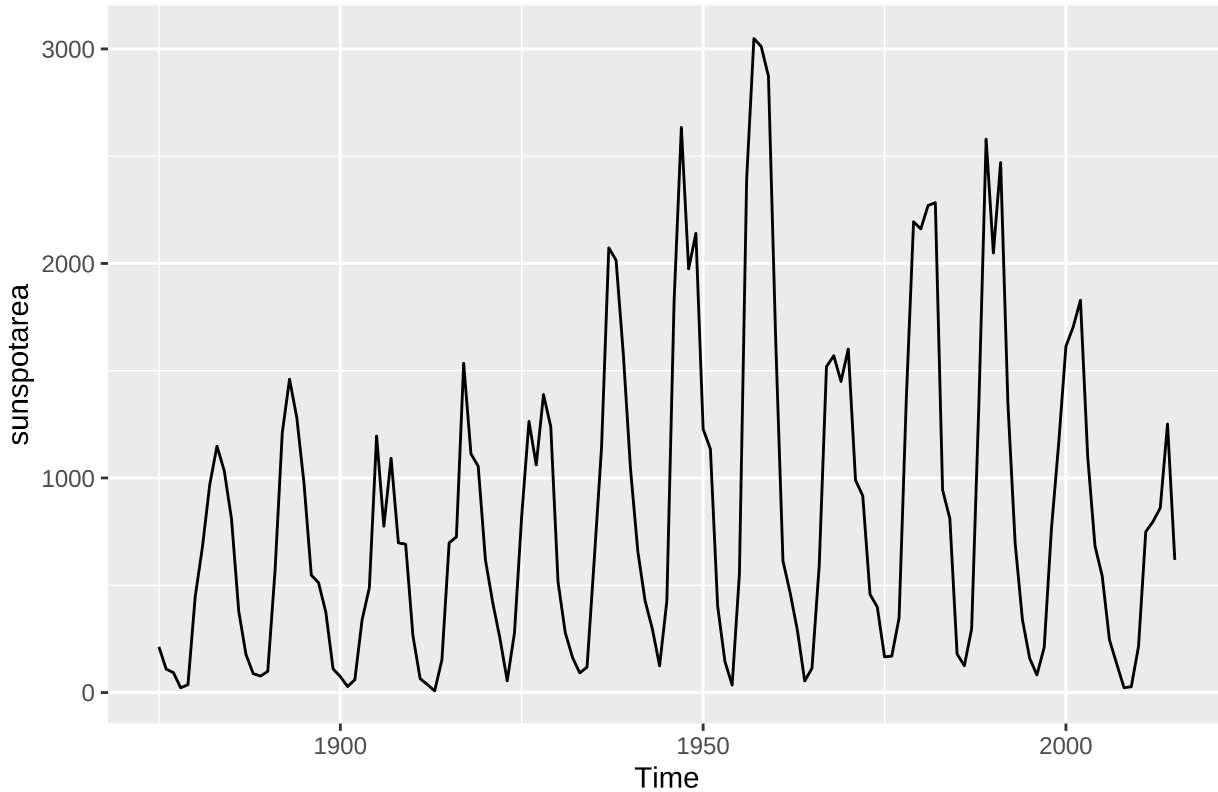
```r
ggAcf(bricksq)
```



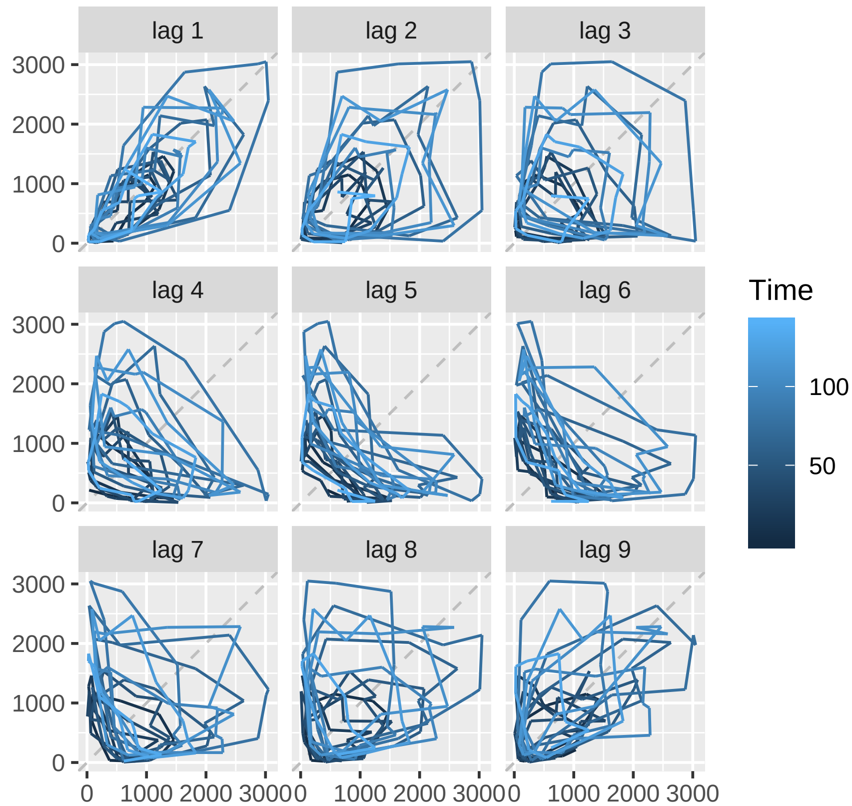
sunspotarea

The data frequency is yearly. While there certainly does not appear to be a trend, there does seem to be a cyclical component that appears to occur almost every 10 years on the micro level. Warning for not seasonal data appears when attempting to run ggseasonplot and ggsubseriesplot.

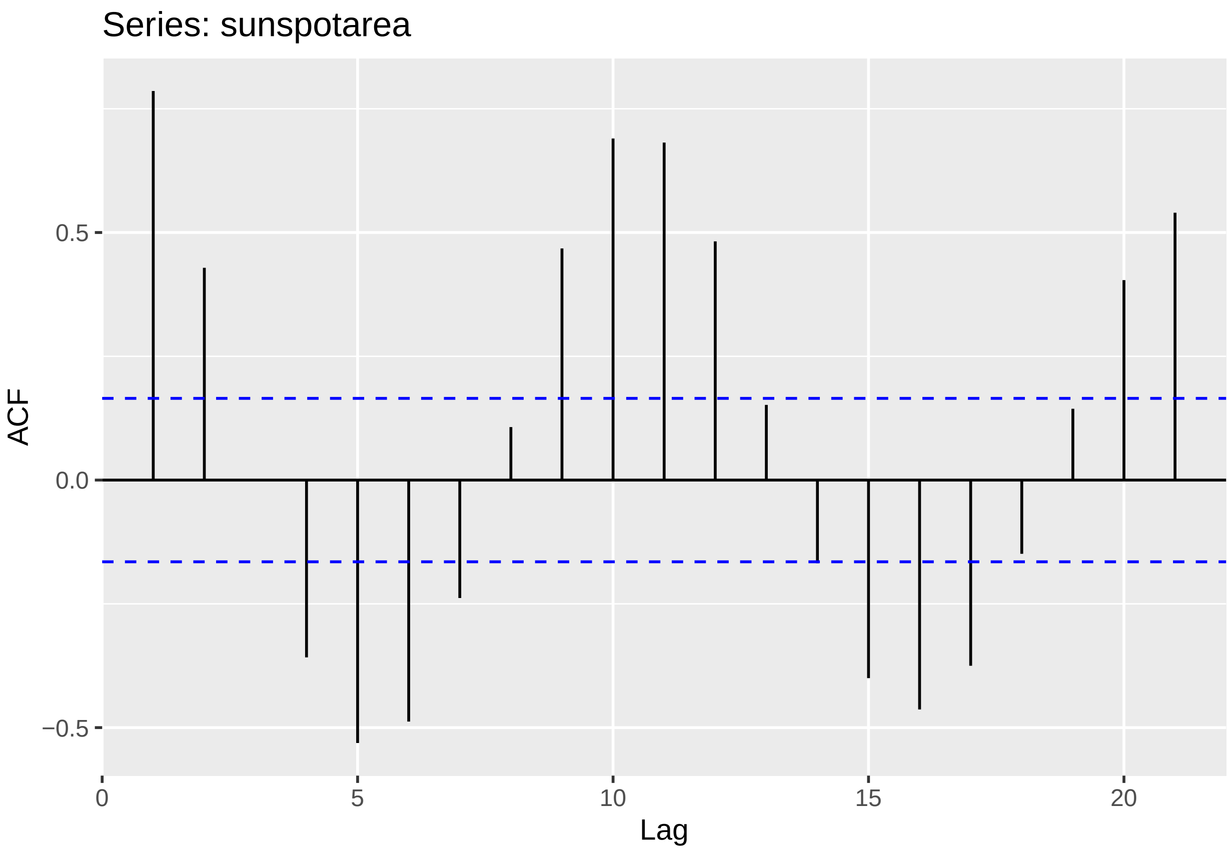
```r
autoplot(sunspotarea)
```




```
##ggseasonplot(sunspotarea) Error: Data not seasonal
##ggsubseriesplot(sunspotarea) Error: Data not seasonal
gglagplot(sunspotarea)
```



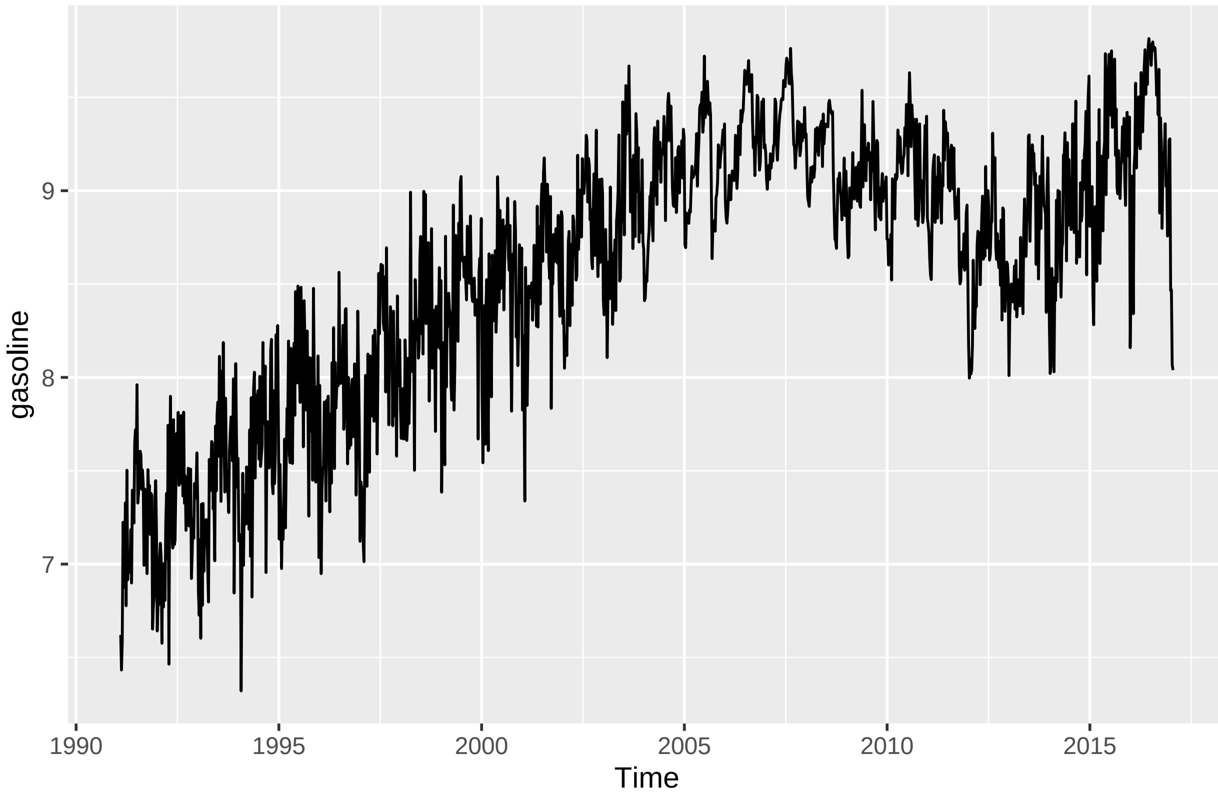

```r
ggAcf(sunspotarea)
```



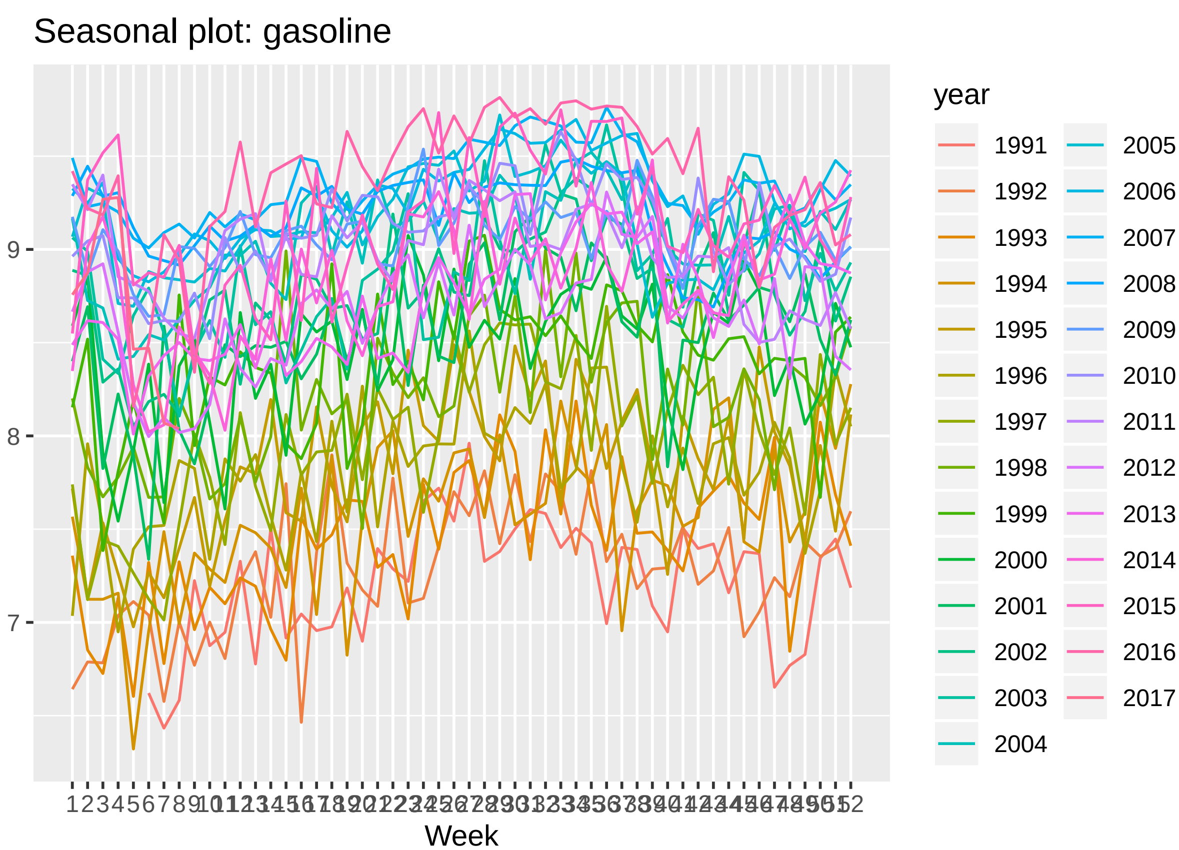
gasoline

Gasoline sales was positively trending with some flattening (or perhaps cycling) towards the later series. Though not obvious cyclical, it is certainly possible that the cycling may occur over the course of multiple decades and that there is not enough evidence to demonstrate it. The lag plot shows positive correlation throughout and the ACF shows some "scalloping" which suggests that there may be some seasonal component as well.

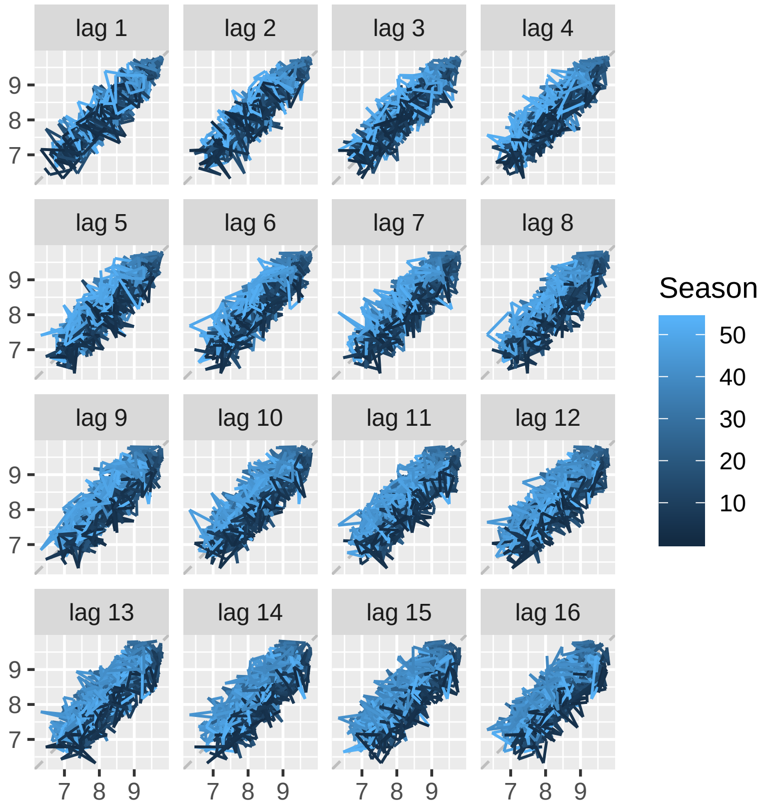
```
autoplot(gasoline)
```



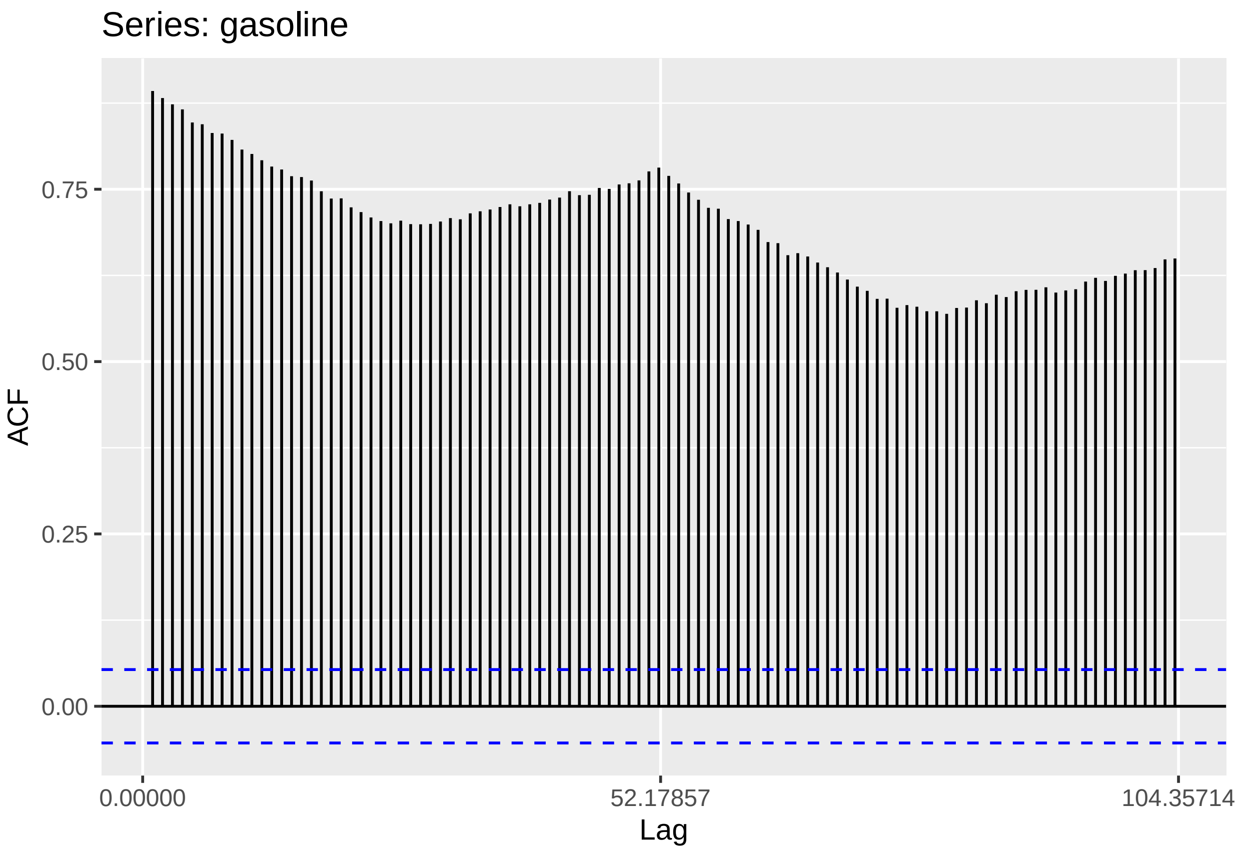

```
ggseasonplot(gasoline)
```



```
##ggsubseriesplot(gasoline) Error: Each season requires at least 2 observations.
gglagplot(gasoline)
```




```
ggAcf(gasoline)
```


Series: gasoline
ACF
0.75
0.50
0.25
0.00
0.00000
52.17857
104.35714
Lag