Homework 1

Forecasting: Principles and Practice - Time Series Graphics

Paul Perez

2/14/2021

```r
library(fpp2)
```

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

## -- Attaching packages ----------------------------------------------------------------- fpp2 2.4 --

## v ggplot2   3.3.3     v fma       2.4
## v forecast  8.13      v expsmooth 2.3

##
```

```r
library(ggplot2)
```

Exercise 2.10 - 1

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

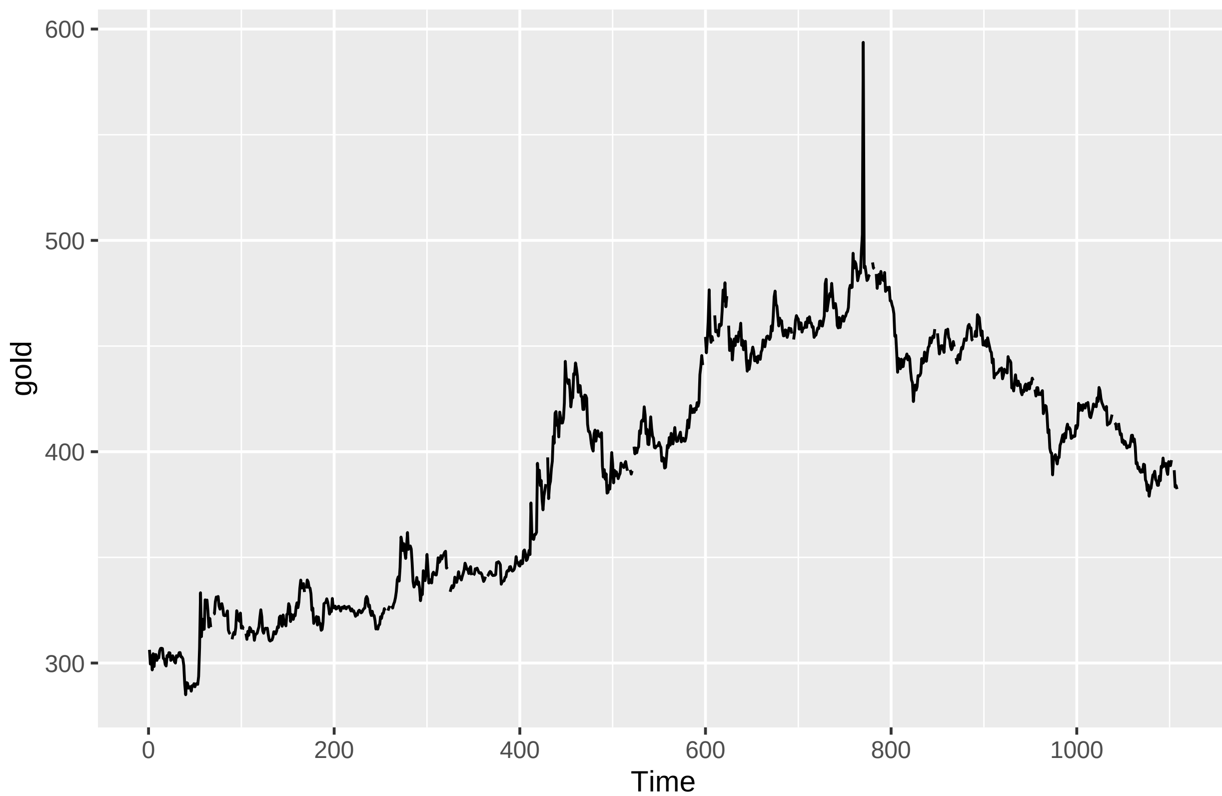
The data in `gold` represents daily morning gold prices in US dollars from January 1st, 1985 - March 31st, 1989.

The data in `woolyrnq` represents quarterly production of woolen yarn in Australia from March 1965 - September 1994.

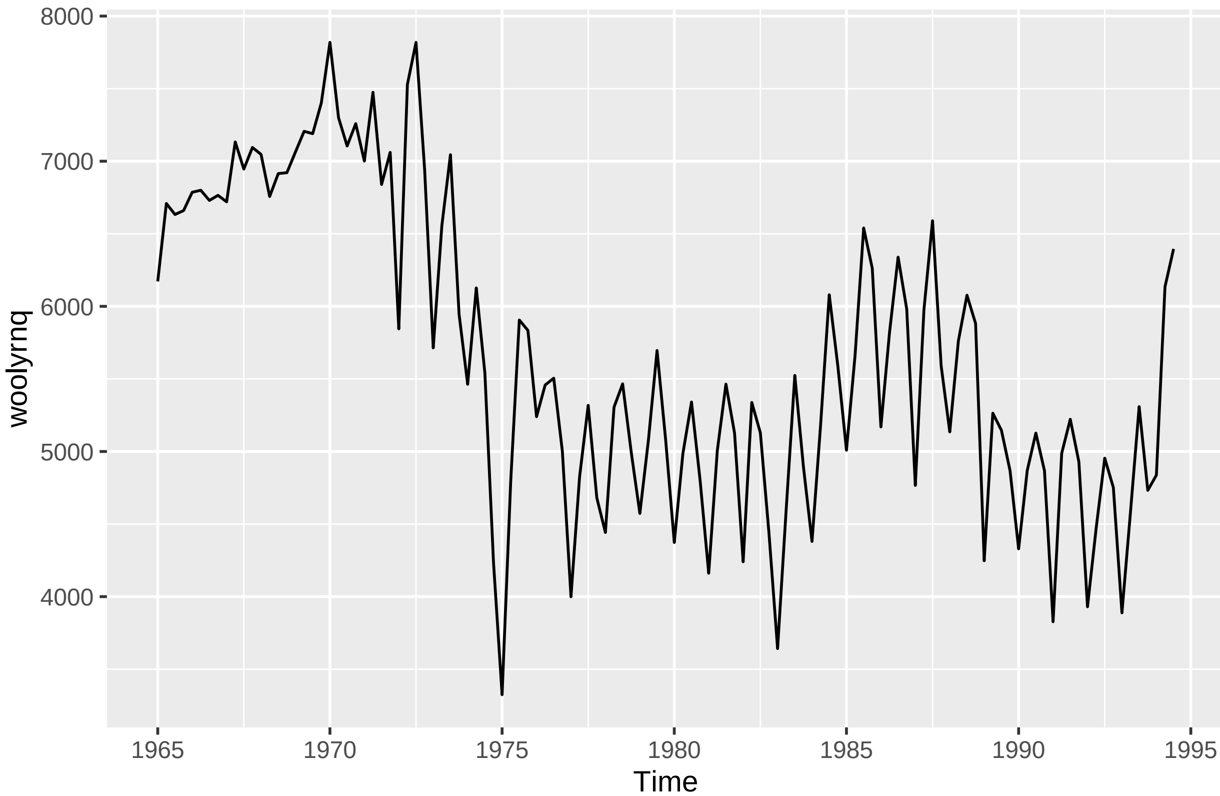
The data in `gas` represents monthly gas production in Australia from 1956 - 1995.

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

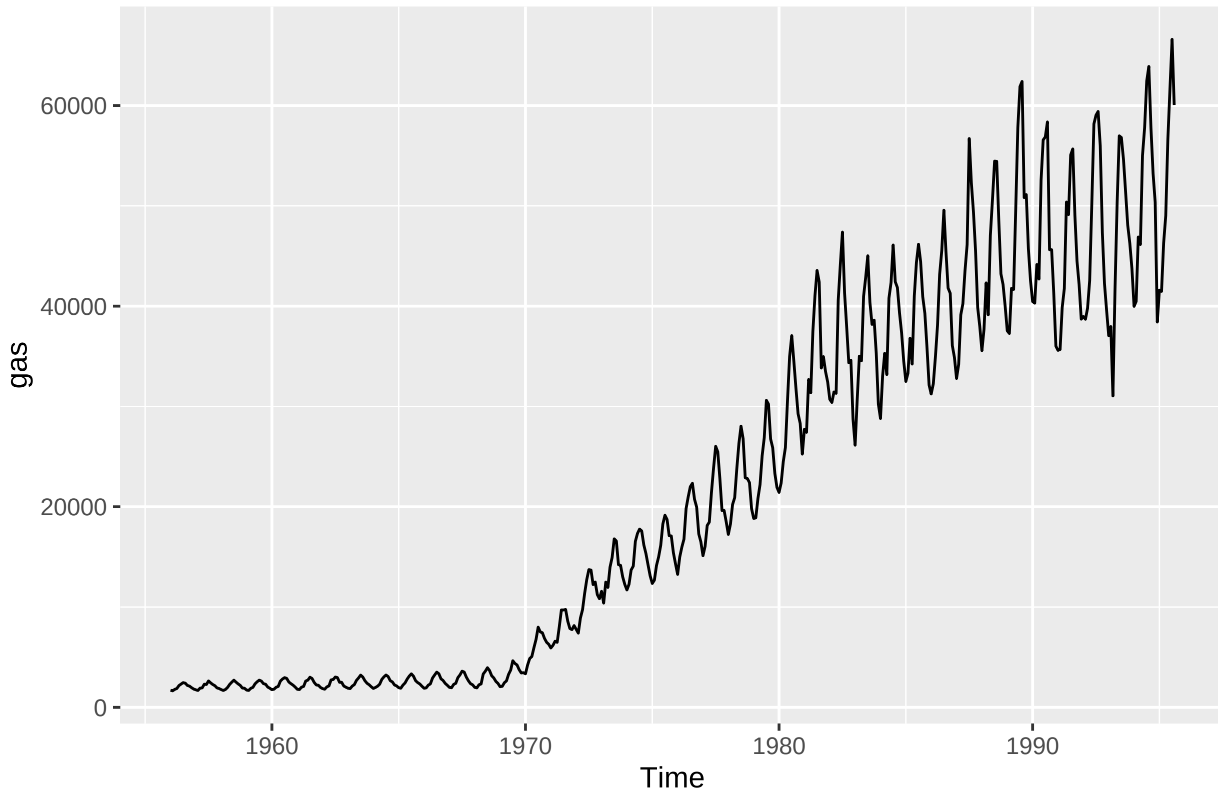
```r
autoplot(gold)
```




```
autoplot(woolyrnq)
```




```
autoplot(gas)
```



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

The frequency for the `gold` series is 1 where the frequency for the `woolyrnq` series is 4, and the frequency for the `gas` series is 12.

```r
frequency(gold)
```

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

```r
frequency(woolyrnq)
```

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

```r
frequency(gas)
```

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

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

The outlier in the `gold` series is the 770th observation.

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

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

Exercise 2.10 - 2

Download the file `tute1.csv` from the book website, open in Excel (or some other spreadsheet application), and review its contents. You should find four columns of information. Columns B through D each contain 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("tute1.csv", header=TRUE)
View(tute1)
```

b.

Convert the data to time series

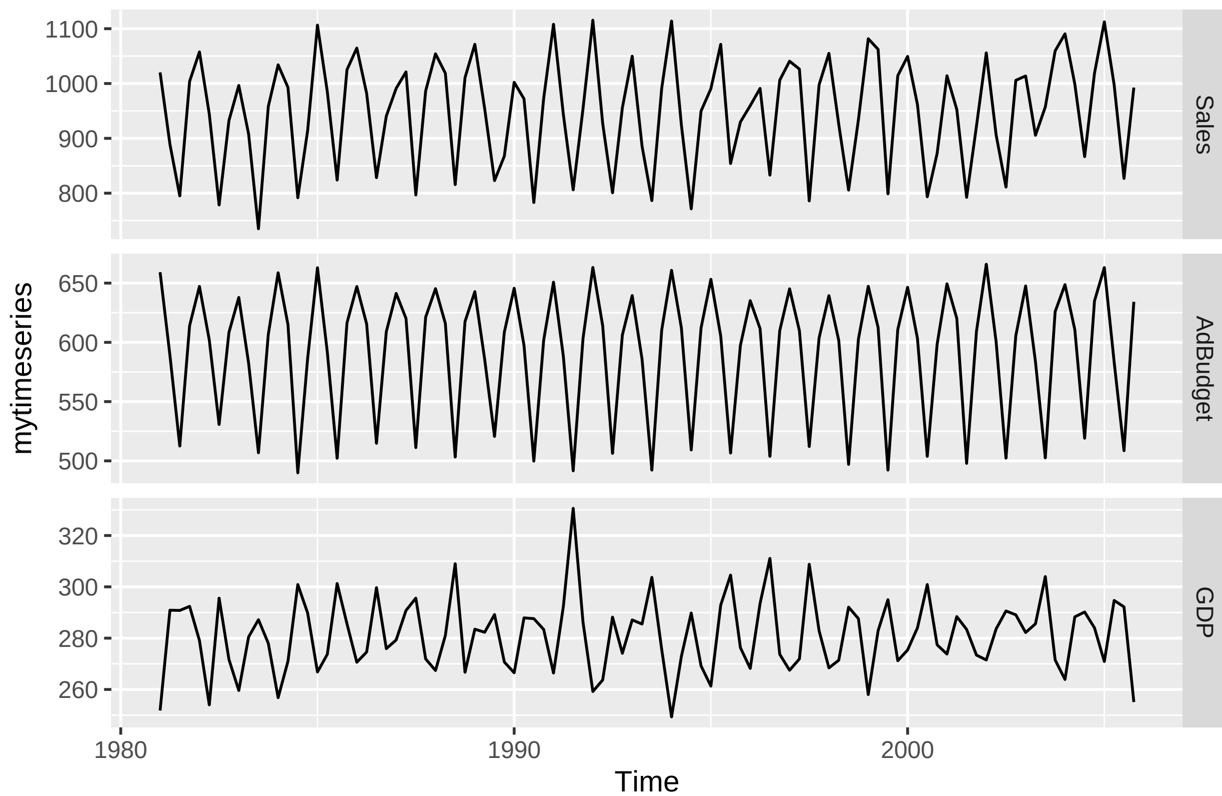
```
mytimeseries <- ts(tute1[,-1], start=1981, frequency=4)
```

(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

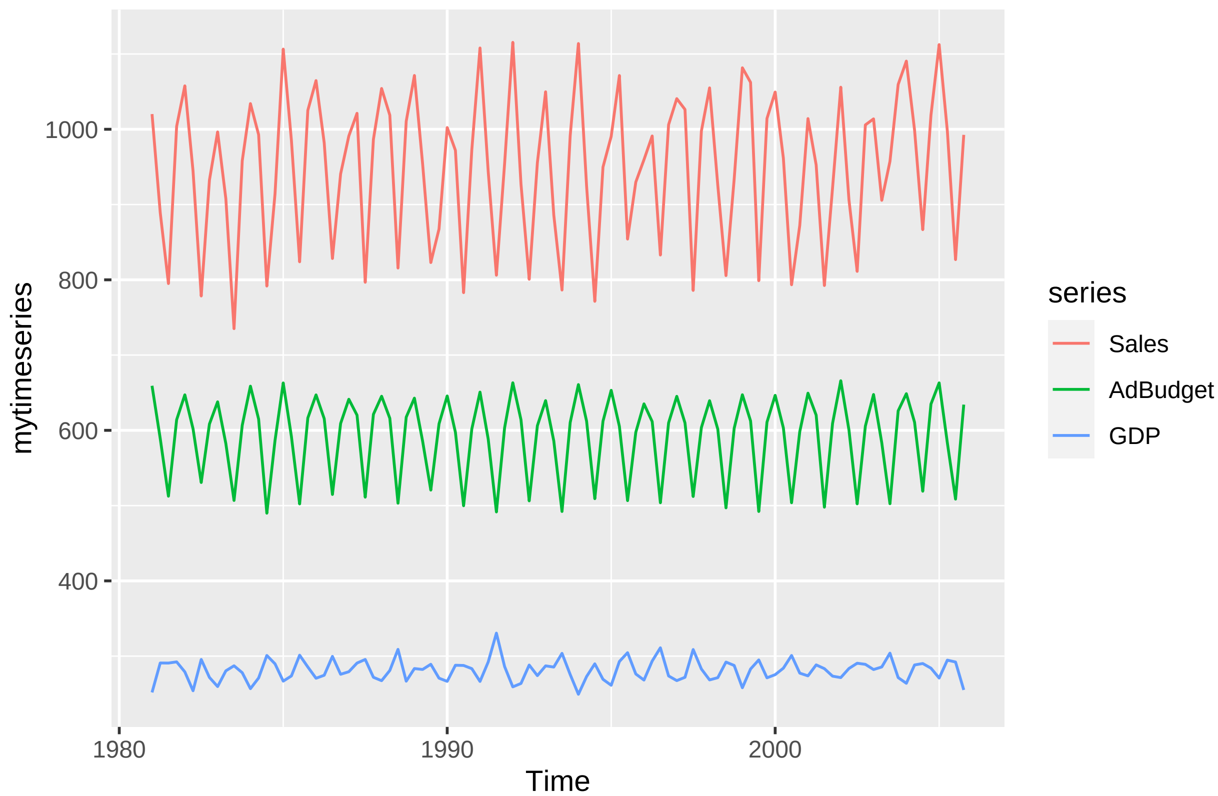
```
autoplot(mytimeseries, facets=TRUE)
```



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

When you exclude `facets=TRUE`, we see each of the series are plotted on the same chart, which differs from the previous example where each series is shown to have its own chart.

```
autoplot(mytimeseries)
```



1000
800
600
400
mytimeseries
1980
1990
2000
Time
series
Sales
AdBudget
GDP

Exercise 2.10 - 3

Download some monthly Australian retail data from the book website. These represent retail sales in variou 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 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):

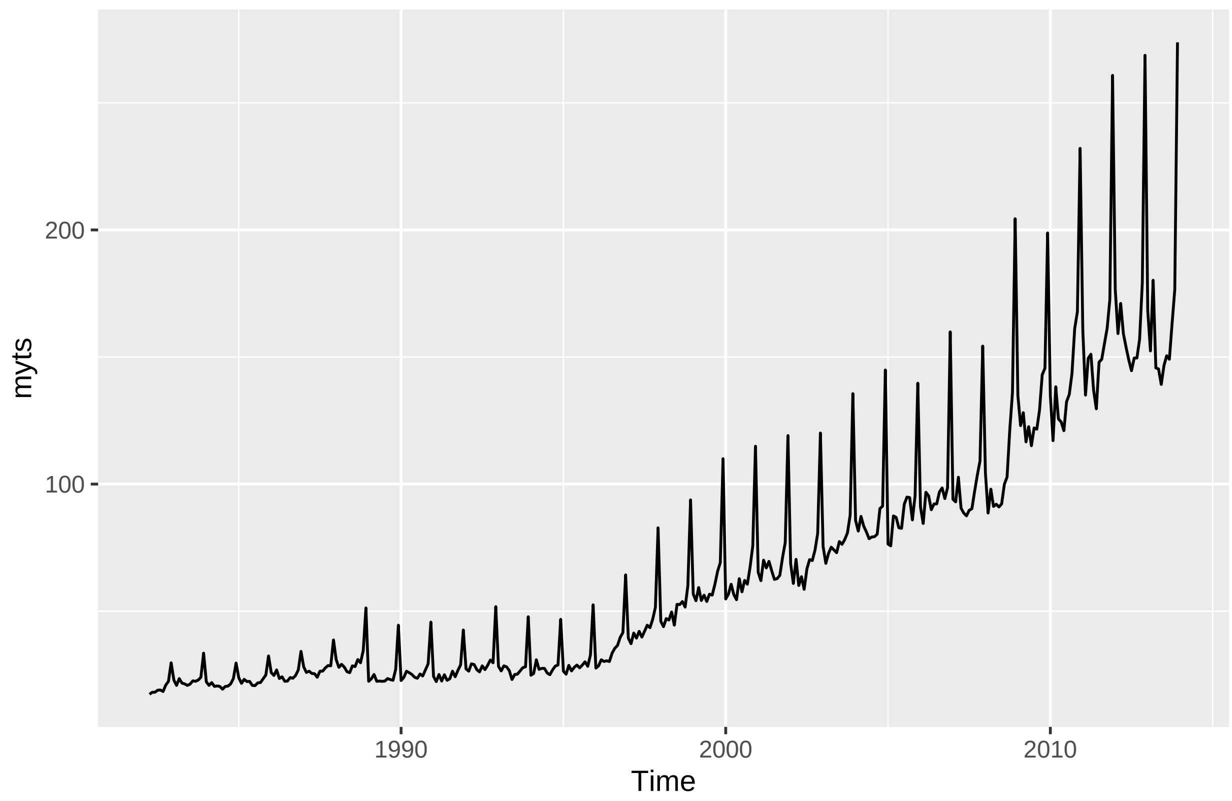
```
myts <- ts(retaildata[,"A3349414R"],
  frequency=12, start=c(1982,4))
```

c.

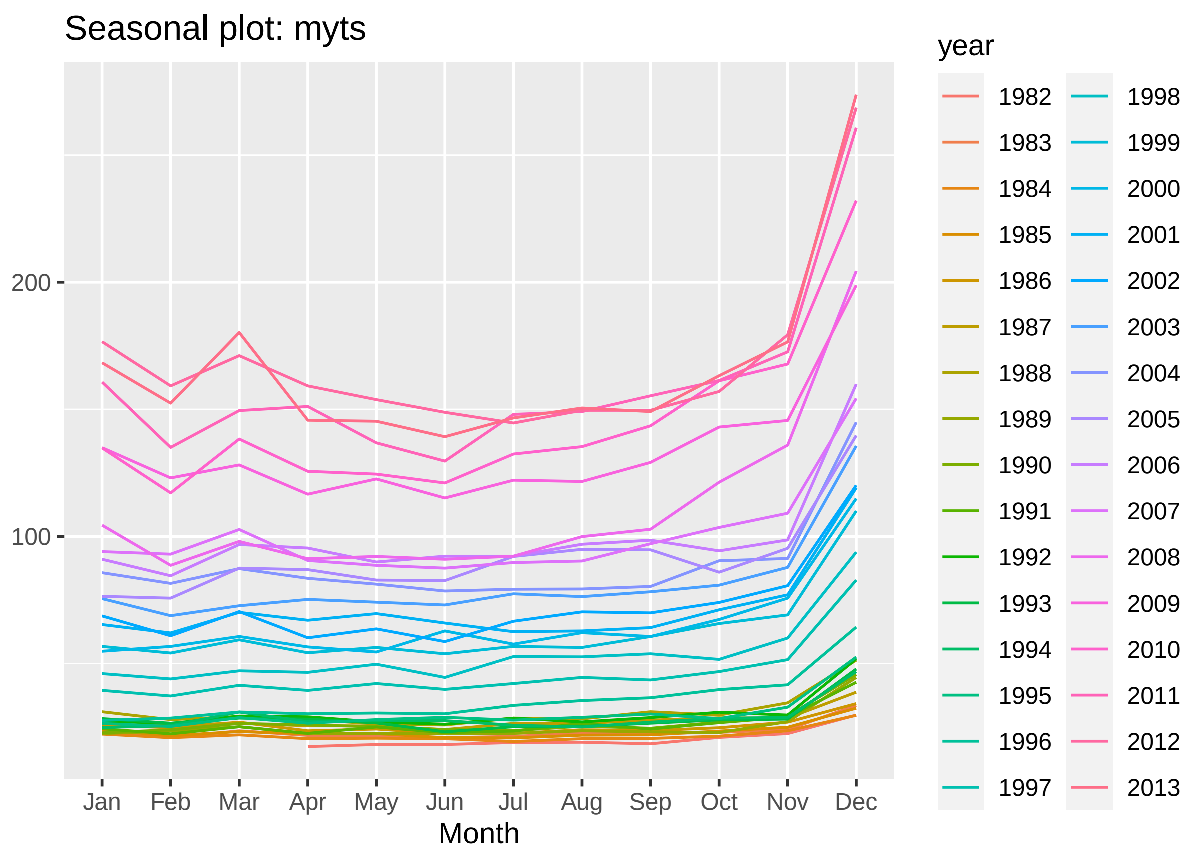
Explore your chosen retail time series using the following functions: `autoplot()`, `ggseasonplot`, `ggsubseriesplot()`, `gglagplot()`, `ggAcf()`

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

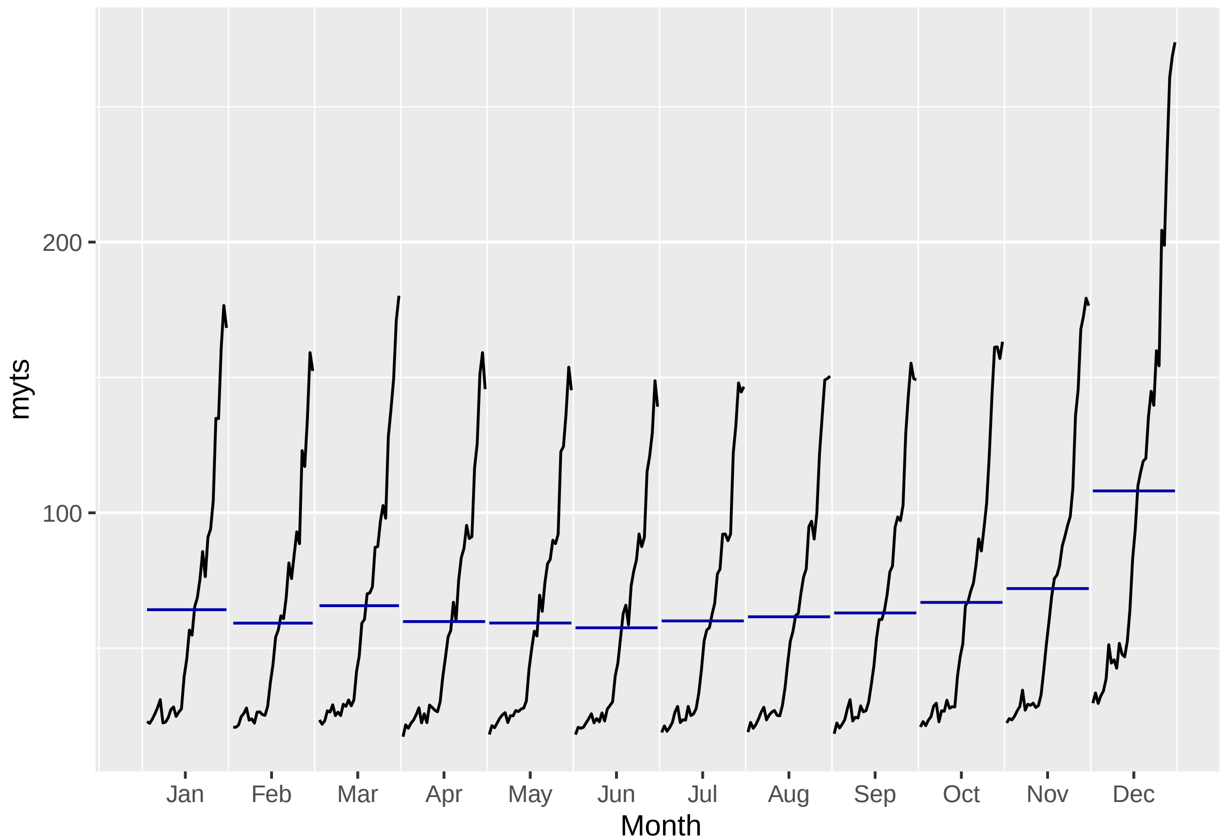
```
autoplot(myts)
```



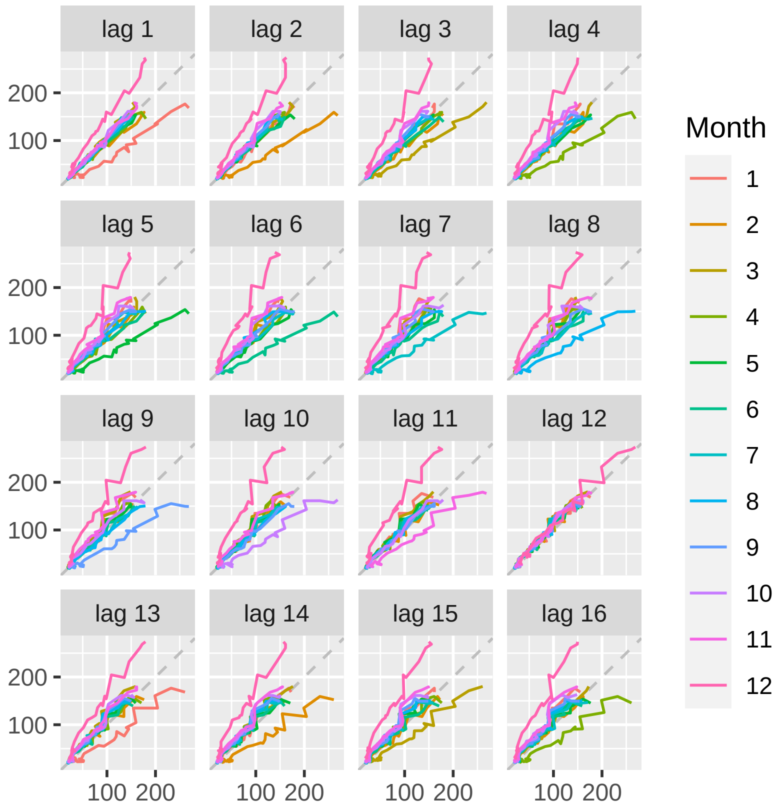

```r
ggseasonplot(myts)
```




```r
ggsubseriesplot(myts)
```

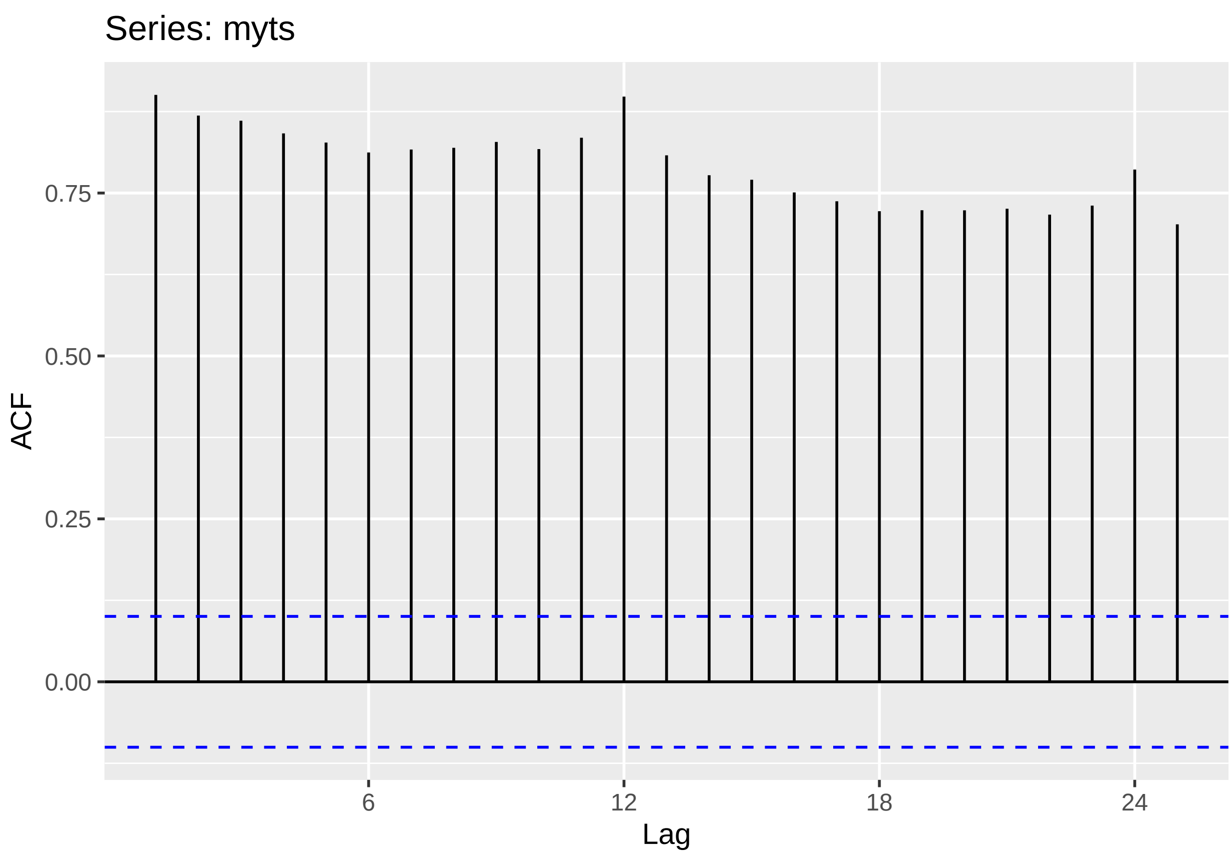



```
gglagplot(myts)
```



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
200
100
100 200
Month
1
2
3
4
5
6
7
8
9
10
11
12


```
ggAcf(myts)
```



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

`A3349414R` represents the column for Liquor retailing in Victoria. Looking at the `autoplot()` function, we can easily identify an upward trend. As we check out the `ggseasonplot()` function, we can see seasonality toward the end of the year, specifically in November and December, which increase in sales are likely due to the holiday's. Using the `ggsubseriesplot()` we can see that the mean of sales for each month is greater in those later months, starting with a gradual increase at the end of the summer but having its greatest increase from November to December.Looking into the `gglagplot()` functions output, we can see that the greatest correlation of lag is displayed with lag 12 meaning that their is a stronger case for seasonality here. To follow, the results of the `ggAcf()` function show that there is a slow decrease in ACF as the lags increase due to a trend, and the peaks of this decrease are for the vales of 12 and 24 showing the correlation of the same period 1 and 2 years back.

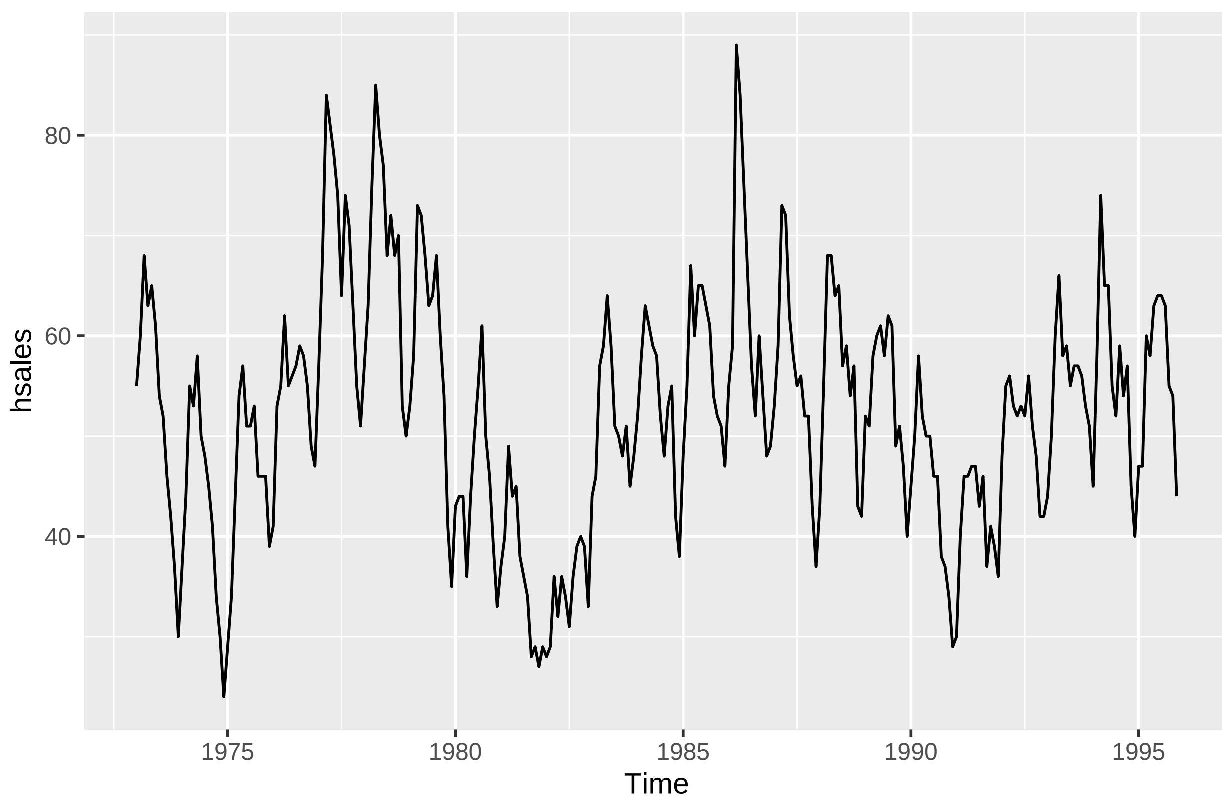
Exercise 2.10 - 6

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

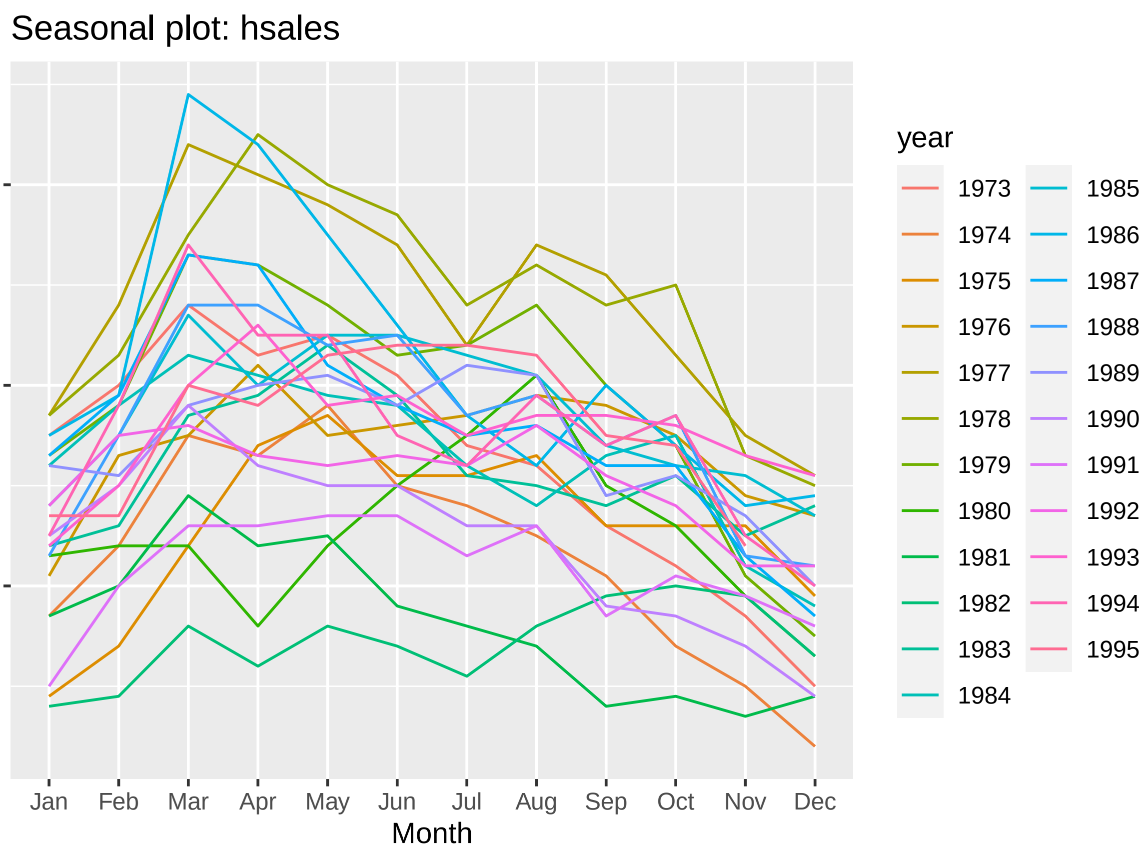
- Can you spot any seasonality, cyclicity and tend?
- What do you learn about the series?

hsales

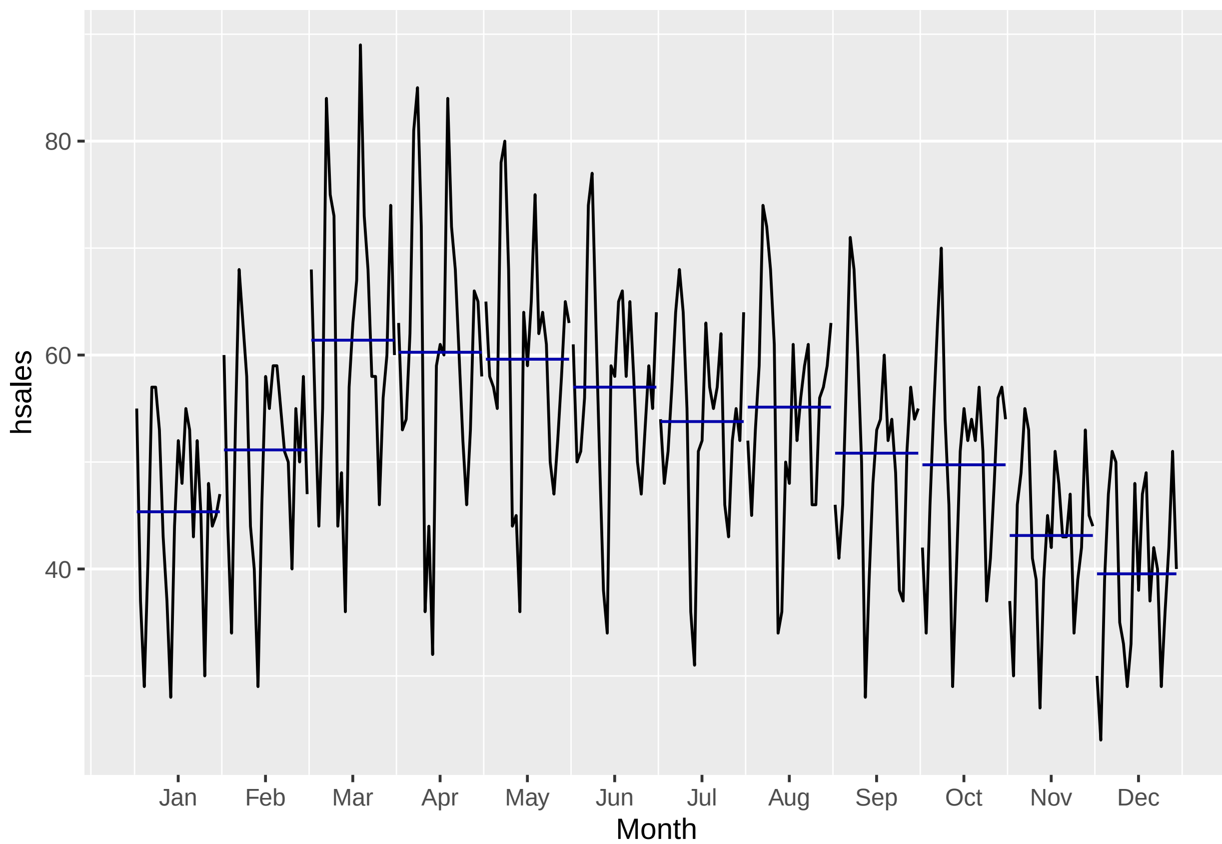
```
autoplot(hsales)
```



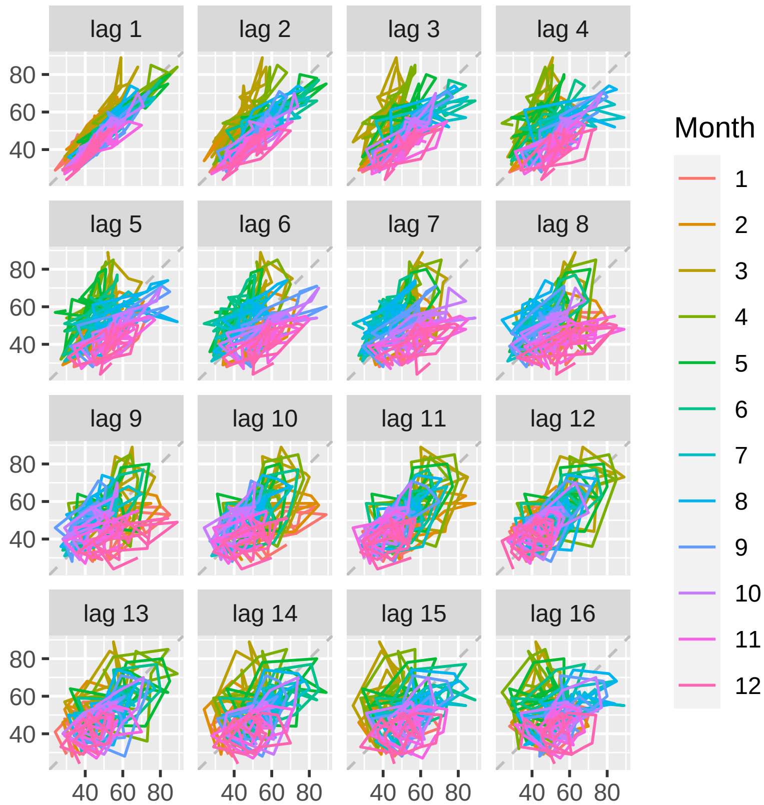
```r
ggseasonplot(hsales)
```



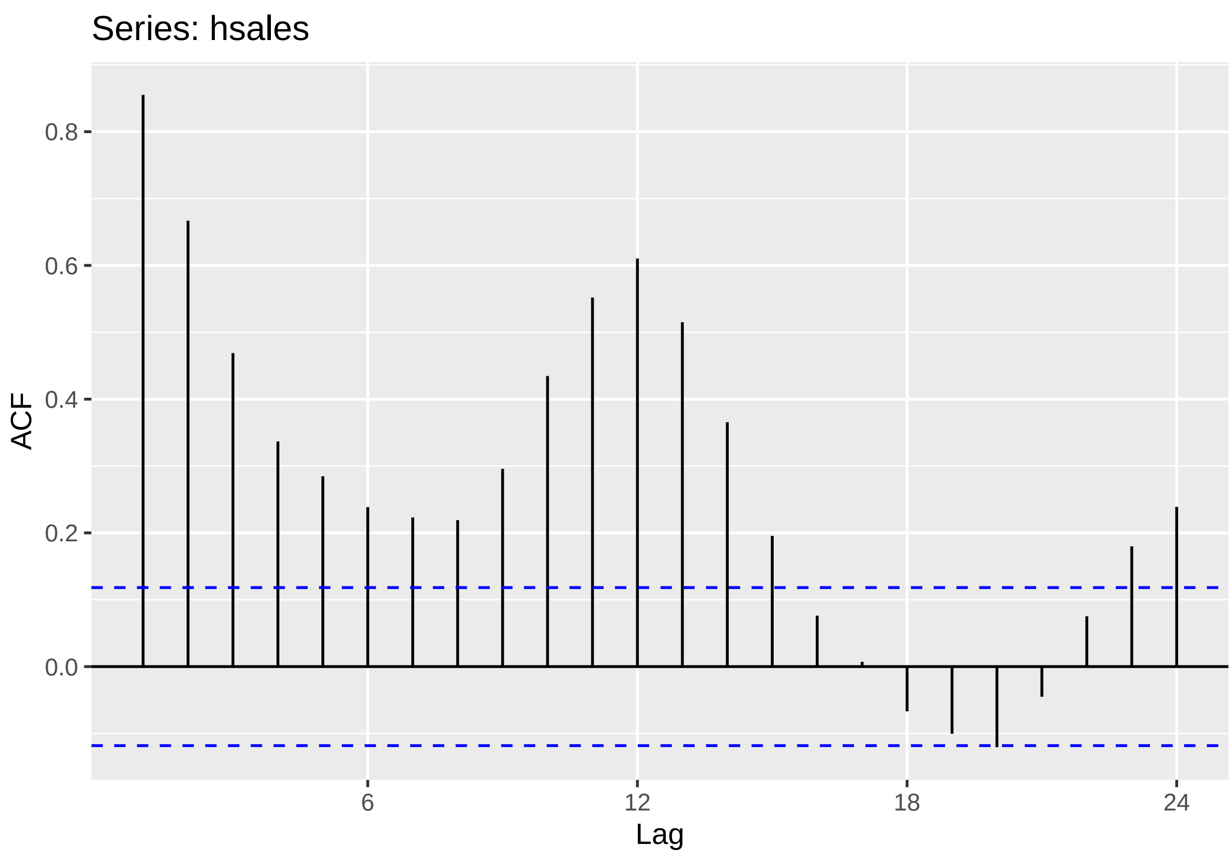

```r
ggsubseriesplot(hsales)
```




```
gglagplot(hsales)
```



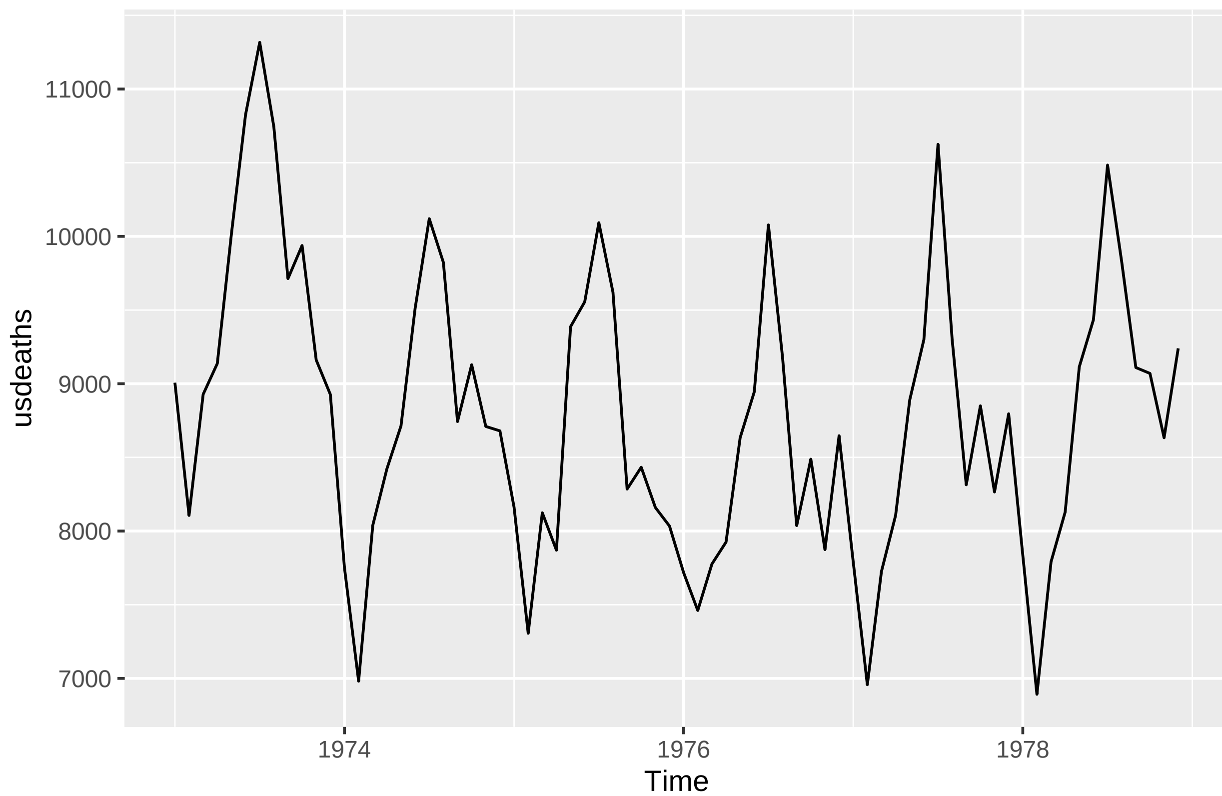

```r
ggAcf(hsales)
```



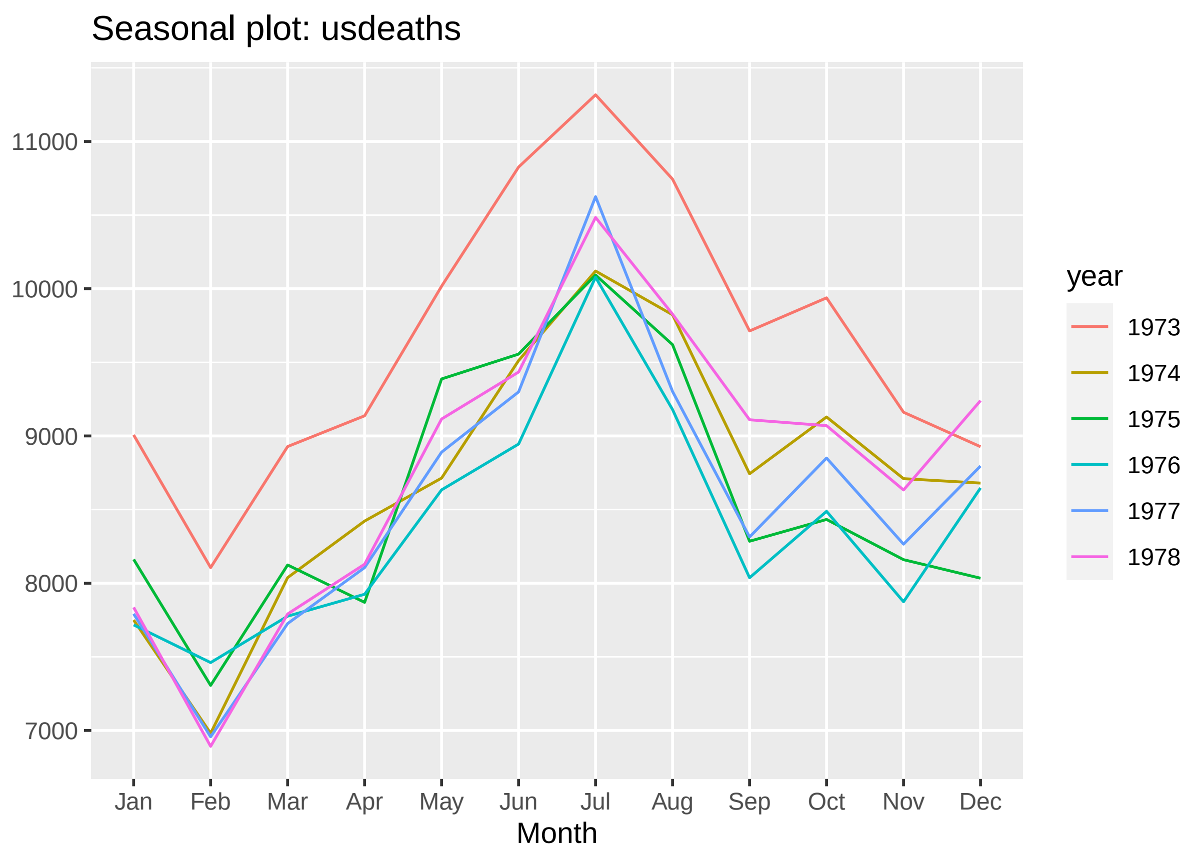
Looking at the `autoplot()` function, we cannot really identify a trend going up or down being that the movement is sideways along the plot. As we check out the `ggseasonplot()` function, we can see seasonality toward the begining of the year, specifically in Q1 of each year. Using the `ggsubseriesplot()` we can see that the mean of sales for each month is greater in those earlier months, starting with a gradual decrease into the summer months.Looking into the `gglagplot()` functions output, we can see that the greatest correlation of lag is displayed with lag 1 and 2 meaning that their is a stronger case for seasonality here. To follow, the results of the `ggAcf()` function show that there is a slow decrease in ACF as the lags increase due to a trend, and the peaks of this decrease are for the vales of 12 and 24 showing the correlation of the same period 1 and 2 years back.

usdeaths

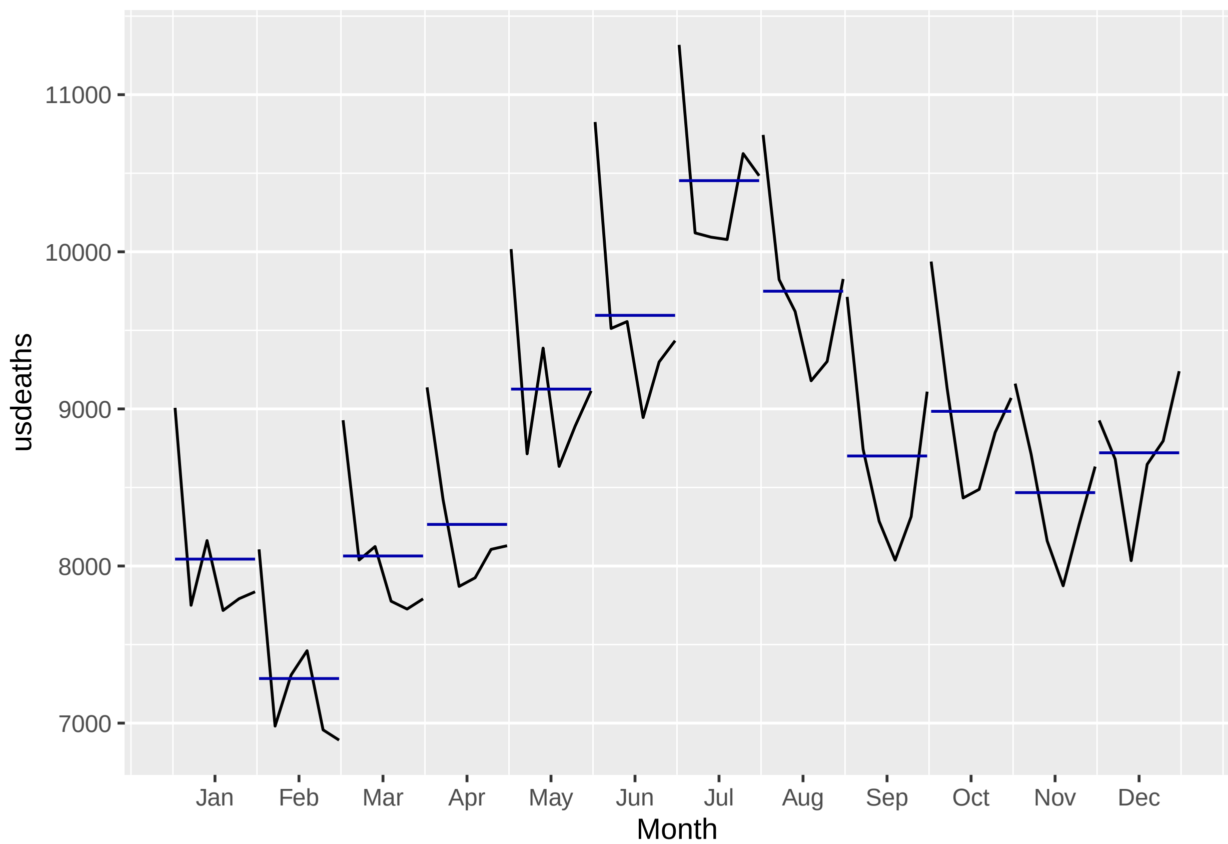
```
autoplot(usdeaths)
```



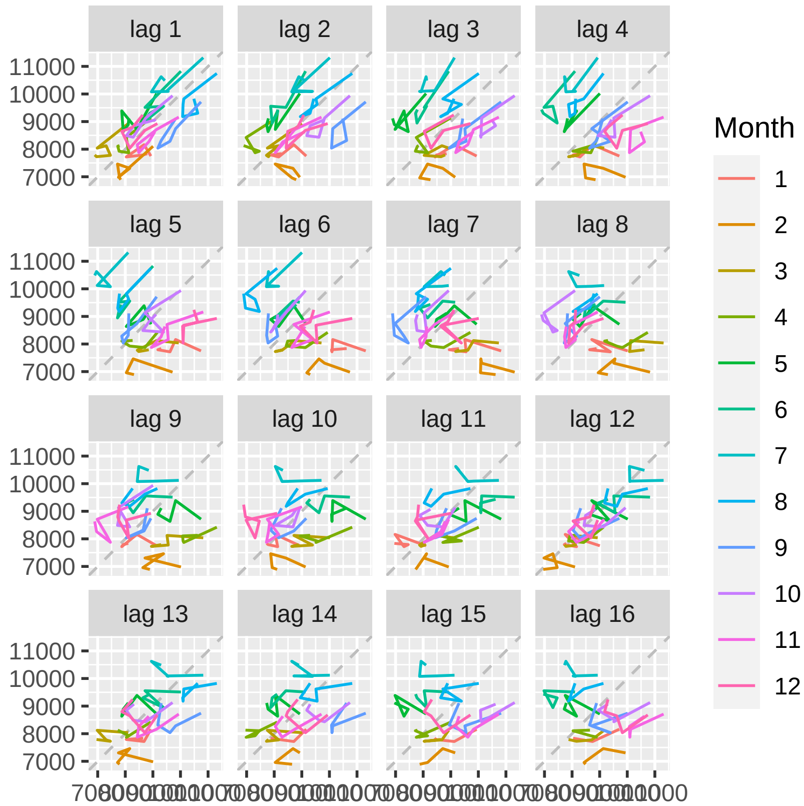

```r
ggseasonplot(usdeaths)
```



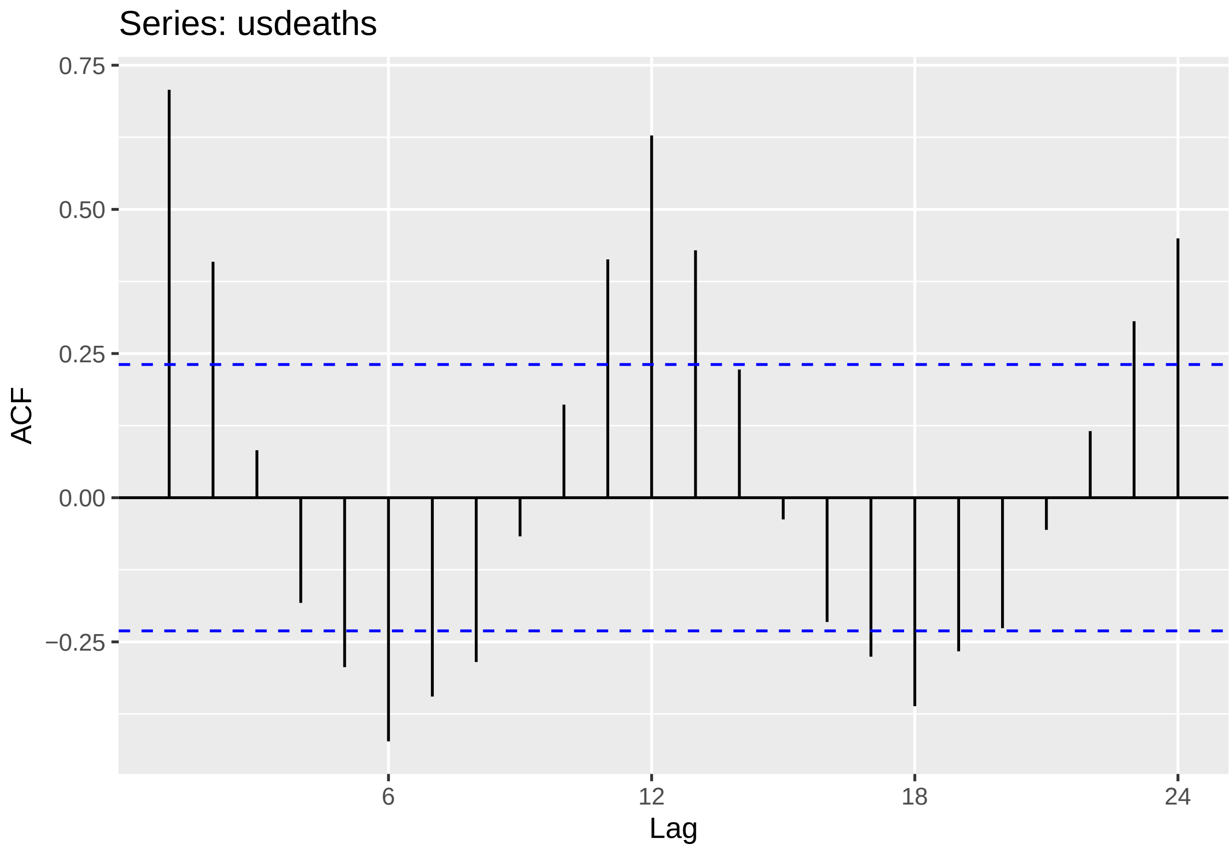

```r
ggsubseriesplot(usdeaths)
```




```r
gglagplot(usdeaths)
```

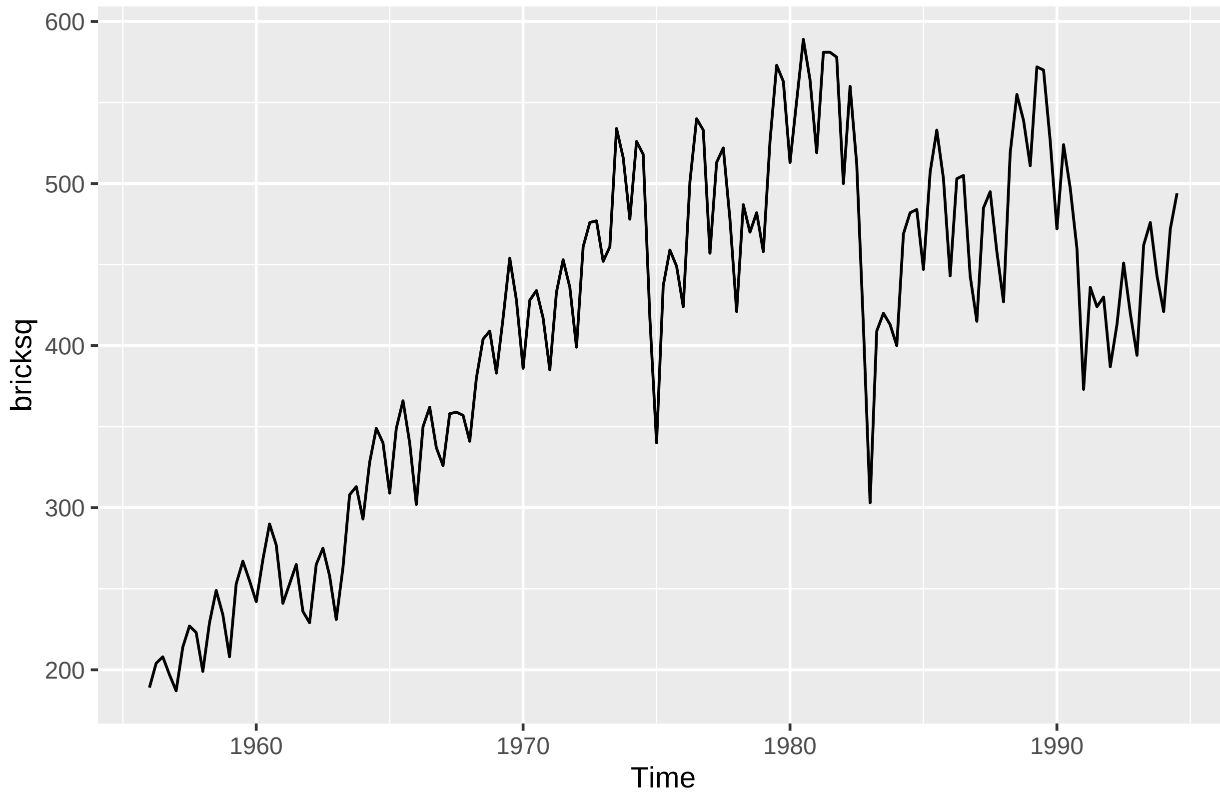



```r
ggAcf(usdeaths)
```

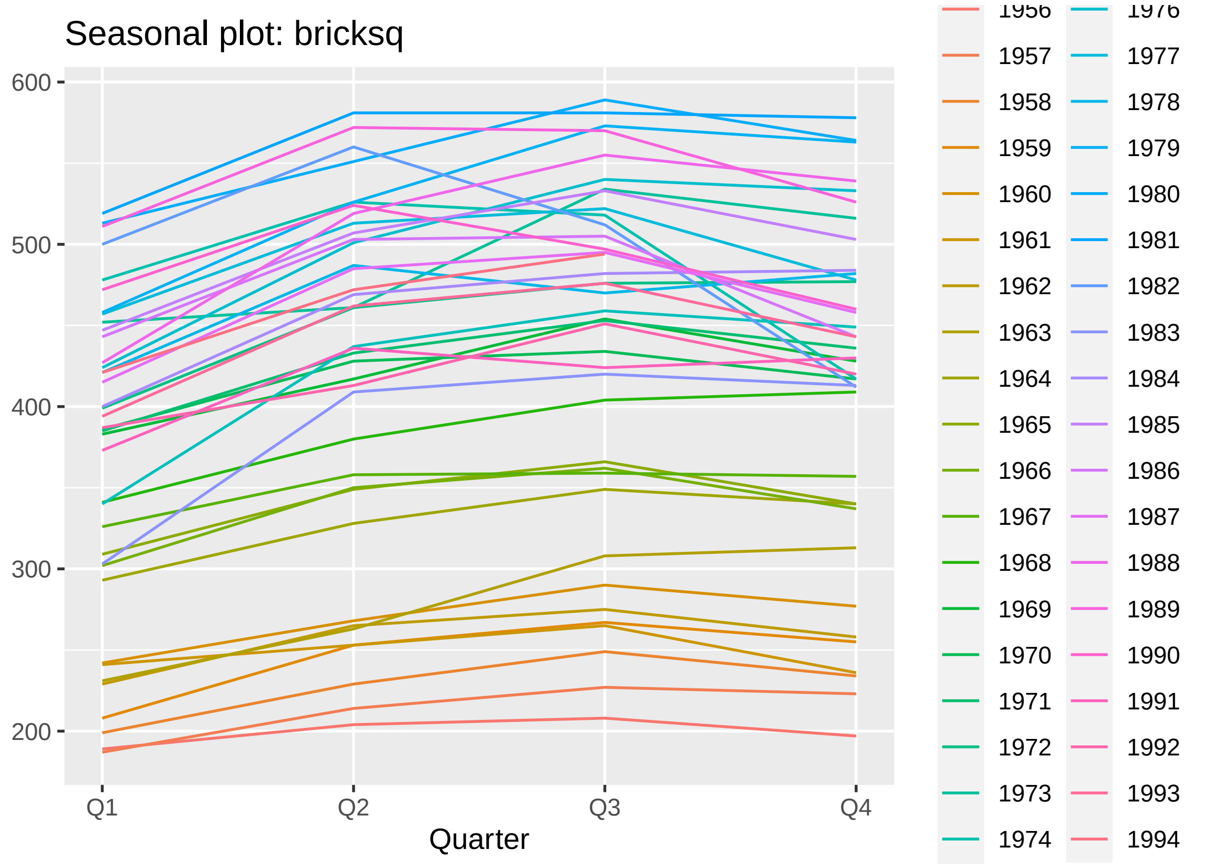


bricksq

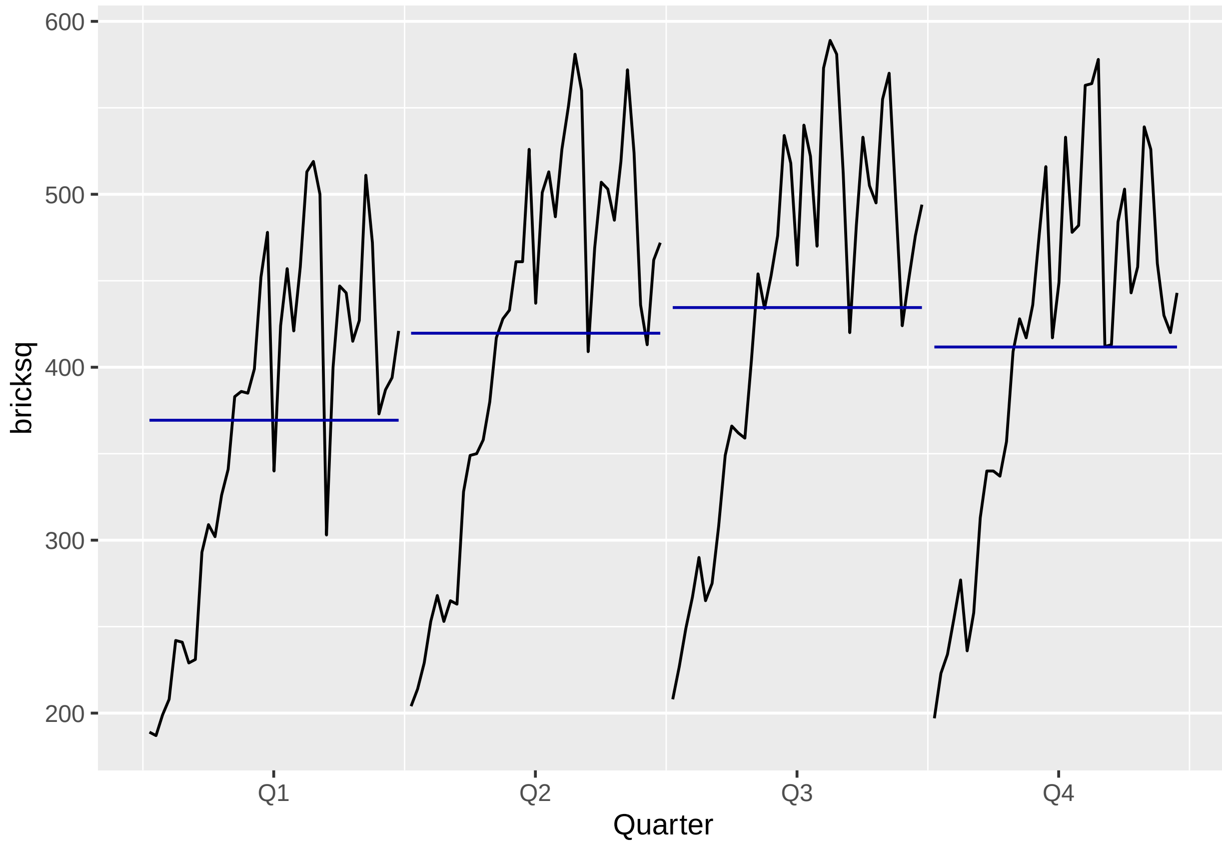
```
autoplot(bricksq)
```



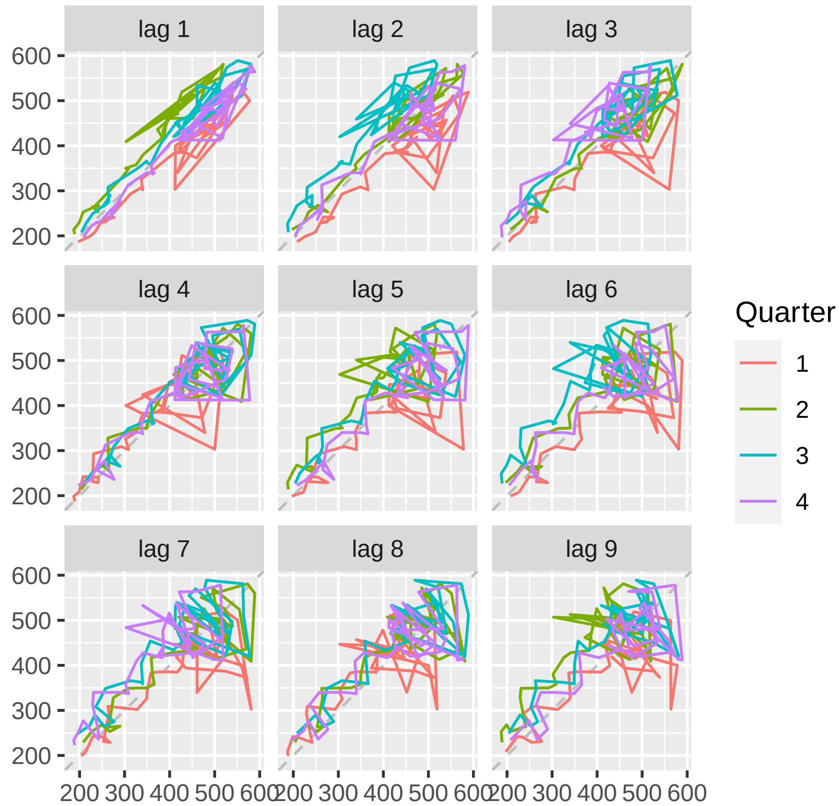

```
ggseasonplot(bricksq)
```



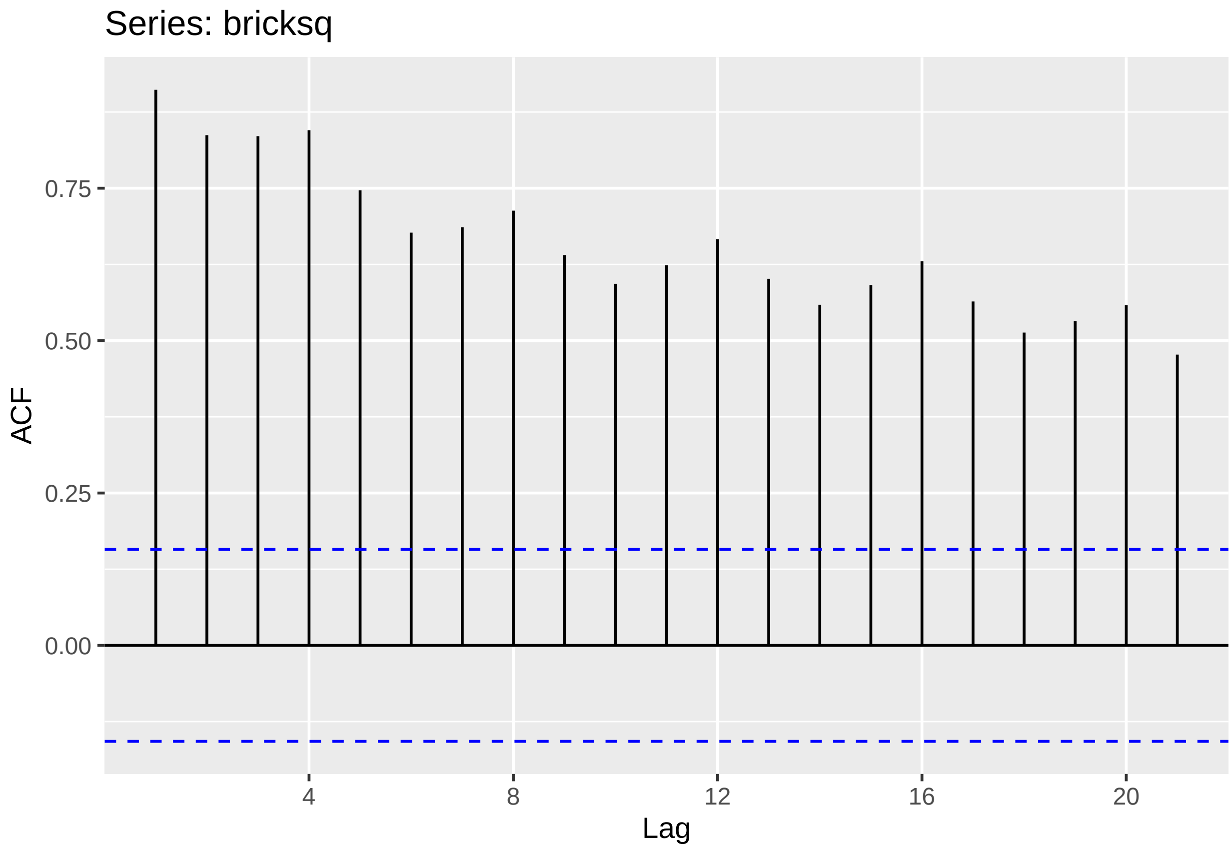

```r
ggsubseriesplot(bricksq)
```




```
gglagplot(bricksq)
```

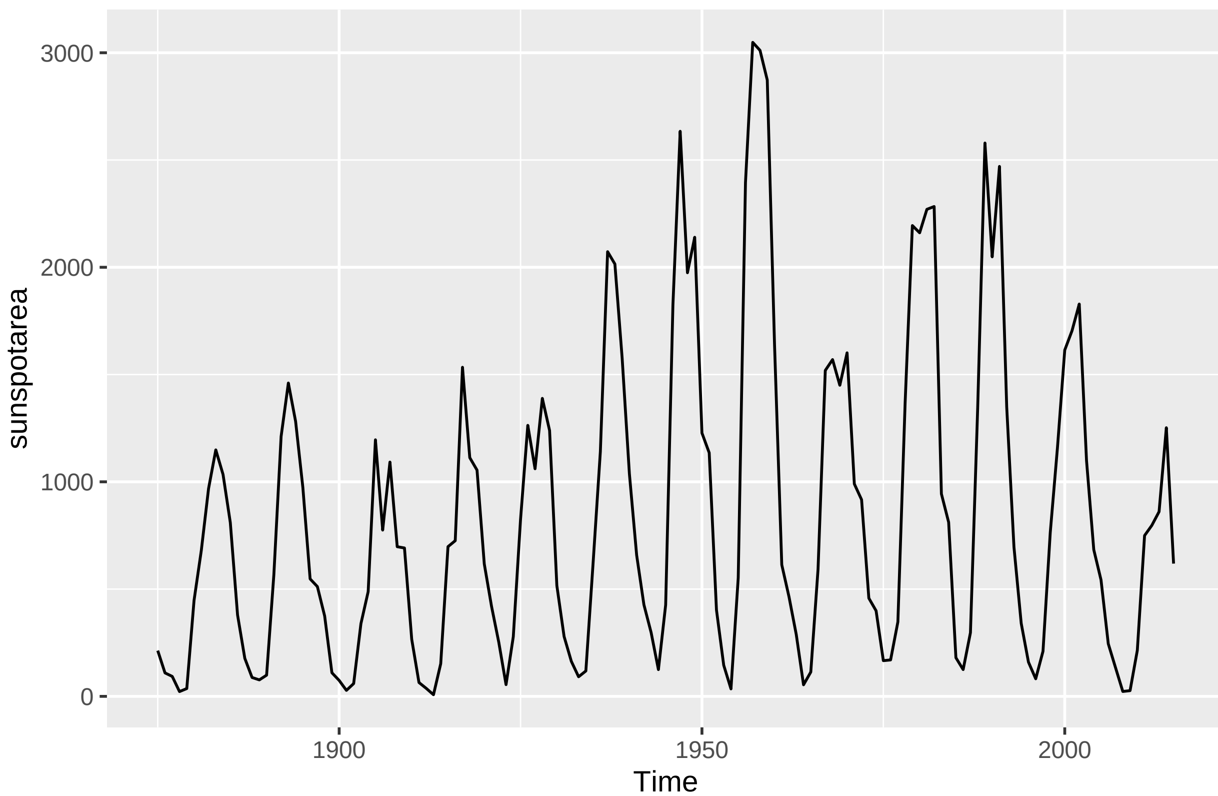



```r
ggAcf(bricksq)
```



sunspotarea

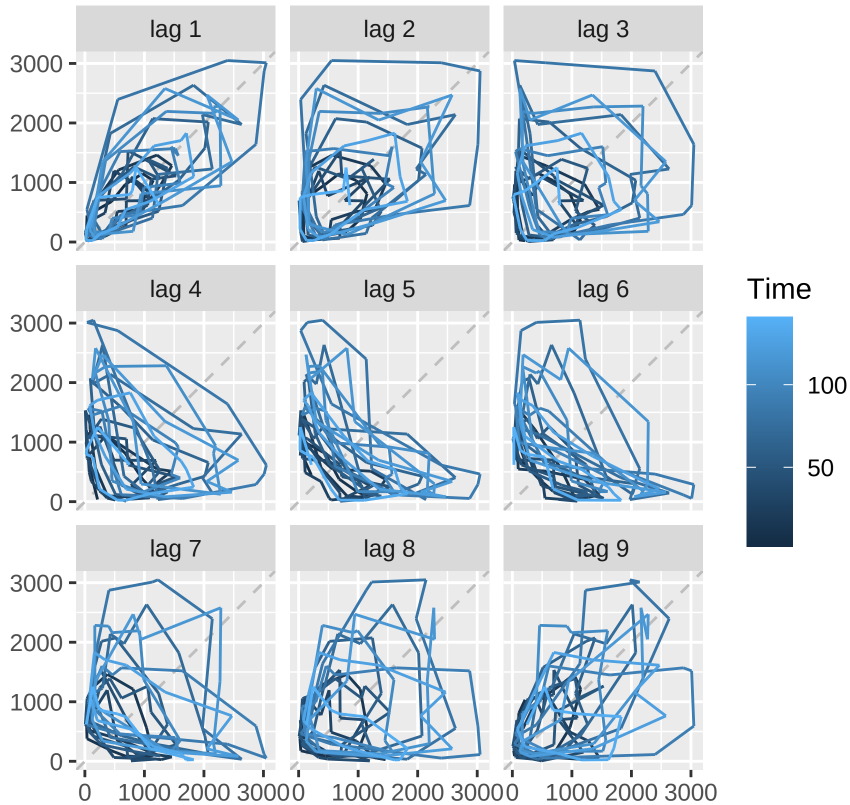
```
autoplot(sunspotarea)
```



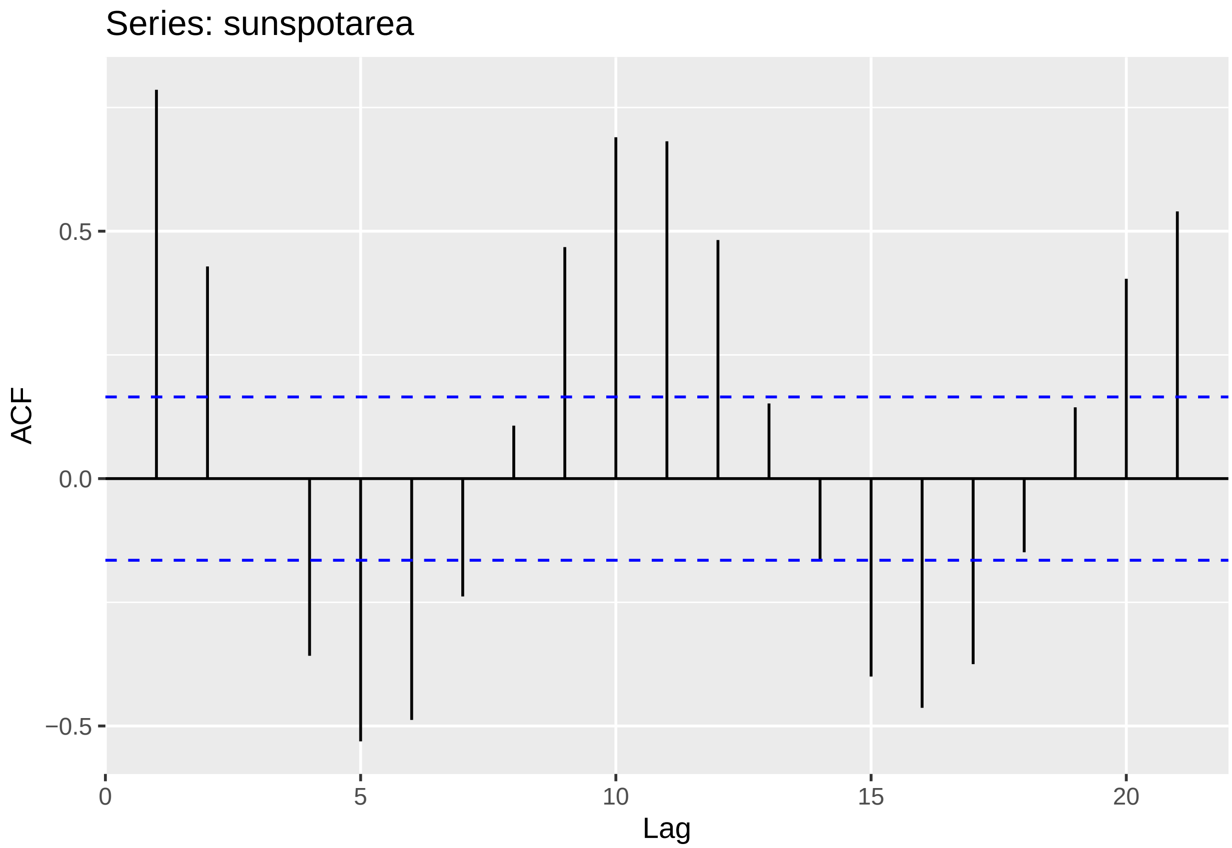

```r
#ggseasonplot(sunspotarea)
```

```r
#ggsubseriesplot(sunspotarea)
```

```r
gglagplot(sunspotarea)
```

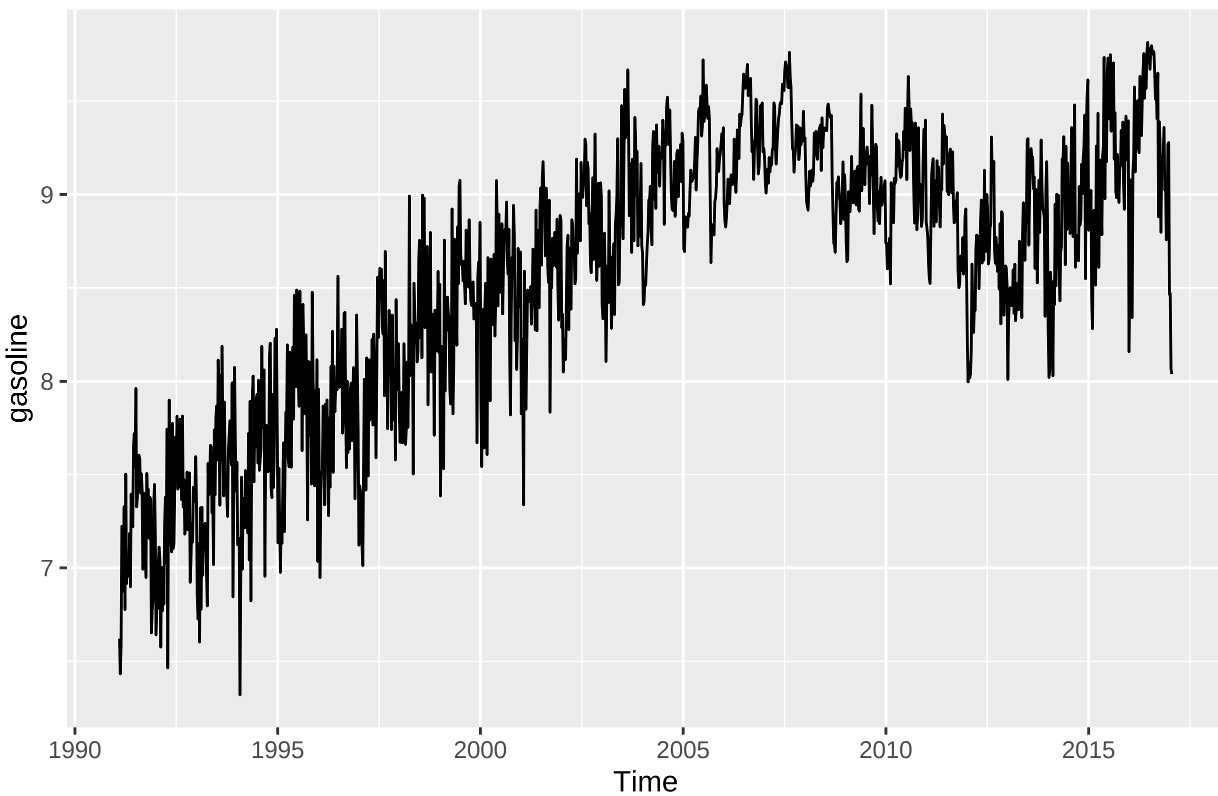



```r
ggAcf(sunspotarea)
```

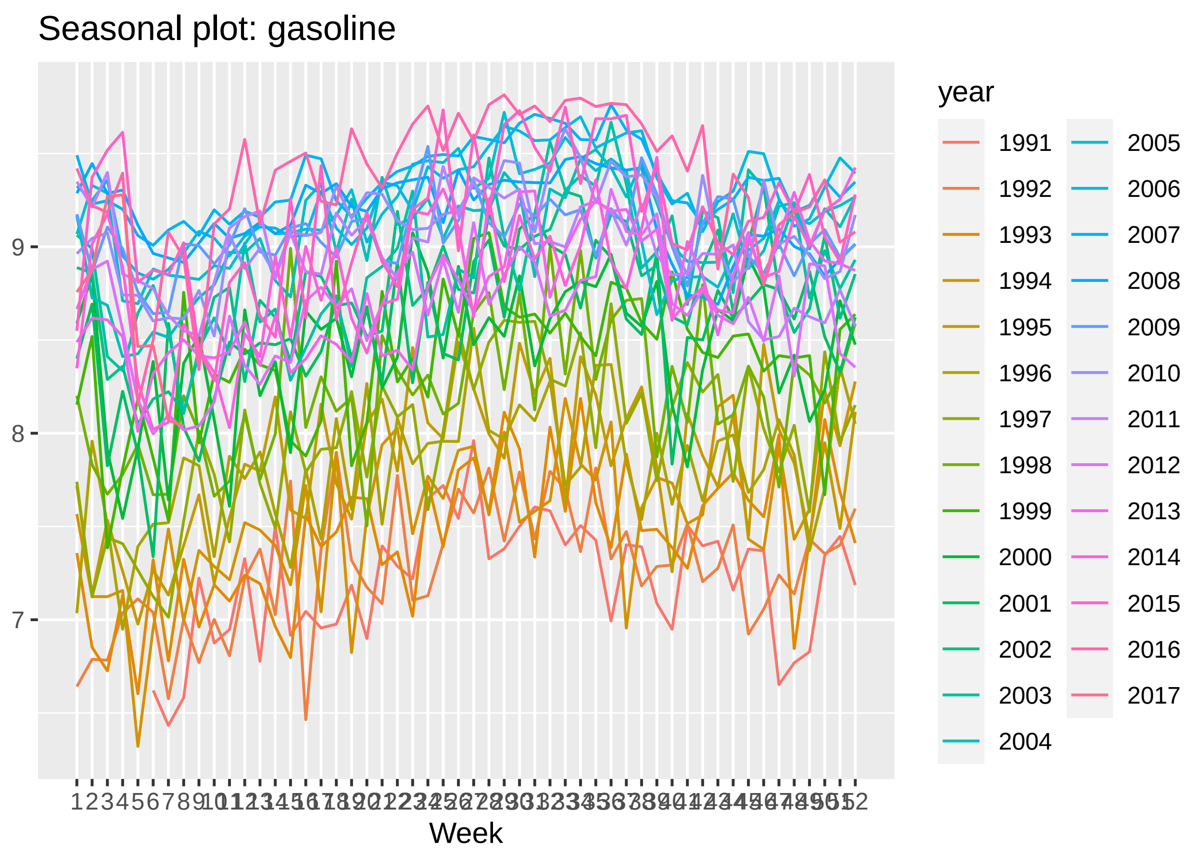


gasoline

```r
autoplot(gasoline)
```

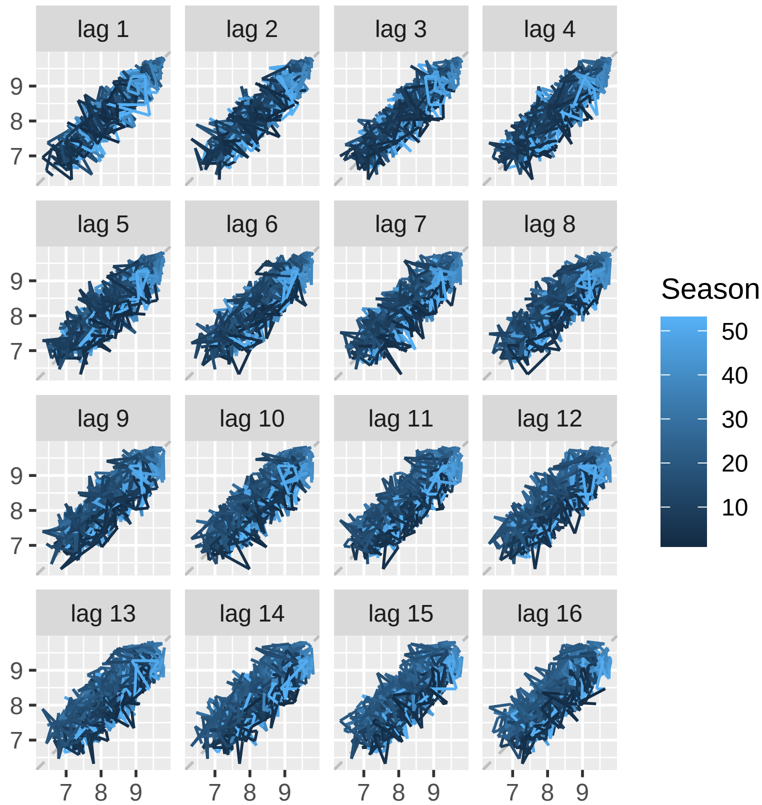



```
ggseasonplot(gasoline)
```

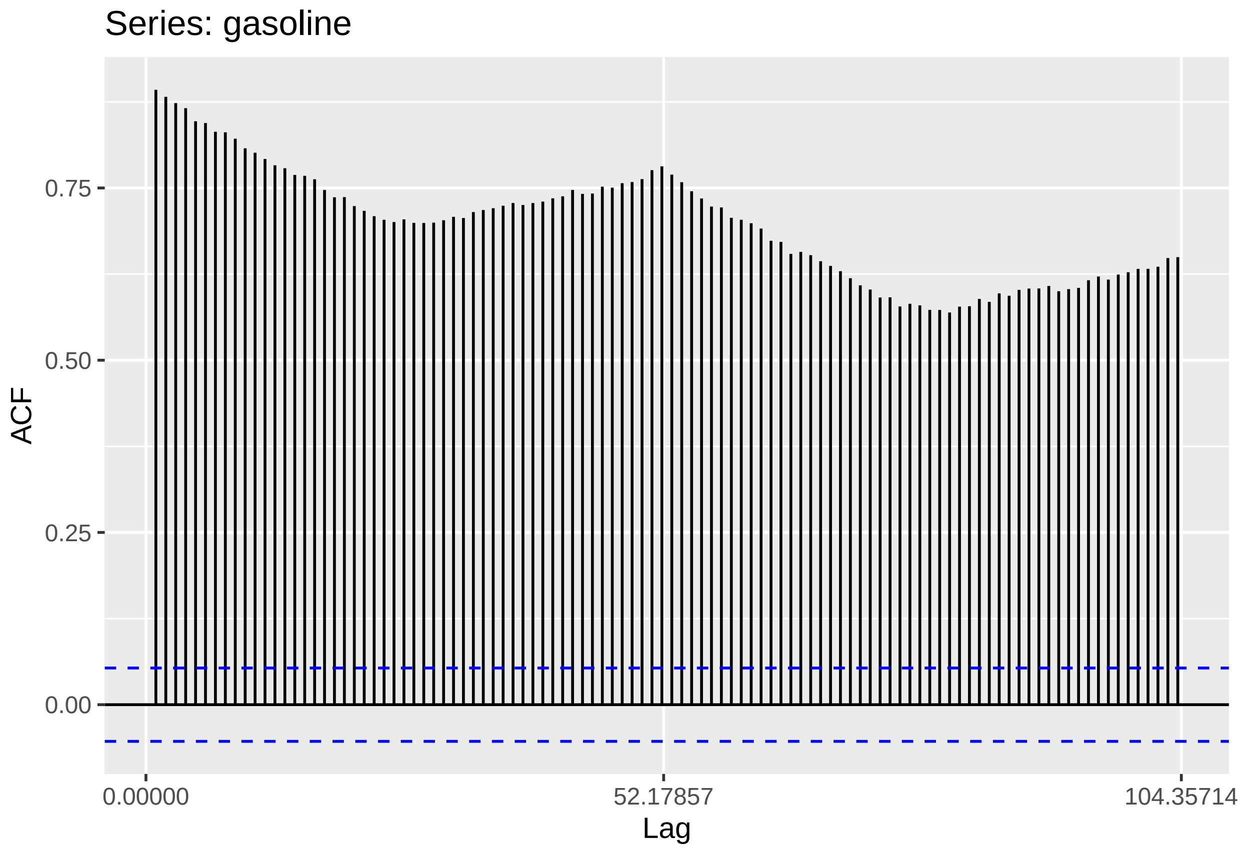



```r
#ggsubseriesplot(gasoline)
```

```r
gglagplot(gasoline)
```




```r
ggAcf(gasoline)
```


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