DATA624: Homework 1

Donald Butler

2023-09-09

Contents

Homework 1

```
library(fpp3)
library(tidyverse)
```

Exercise 2.1

Explore the following four time series: *Bricks* from *aus_production*, *Lynx* from *pelt*, *Close* from *gafa_stock*, *Demand* from *vic_elec*.

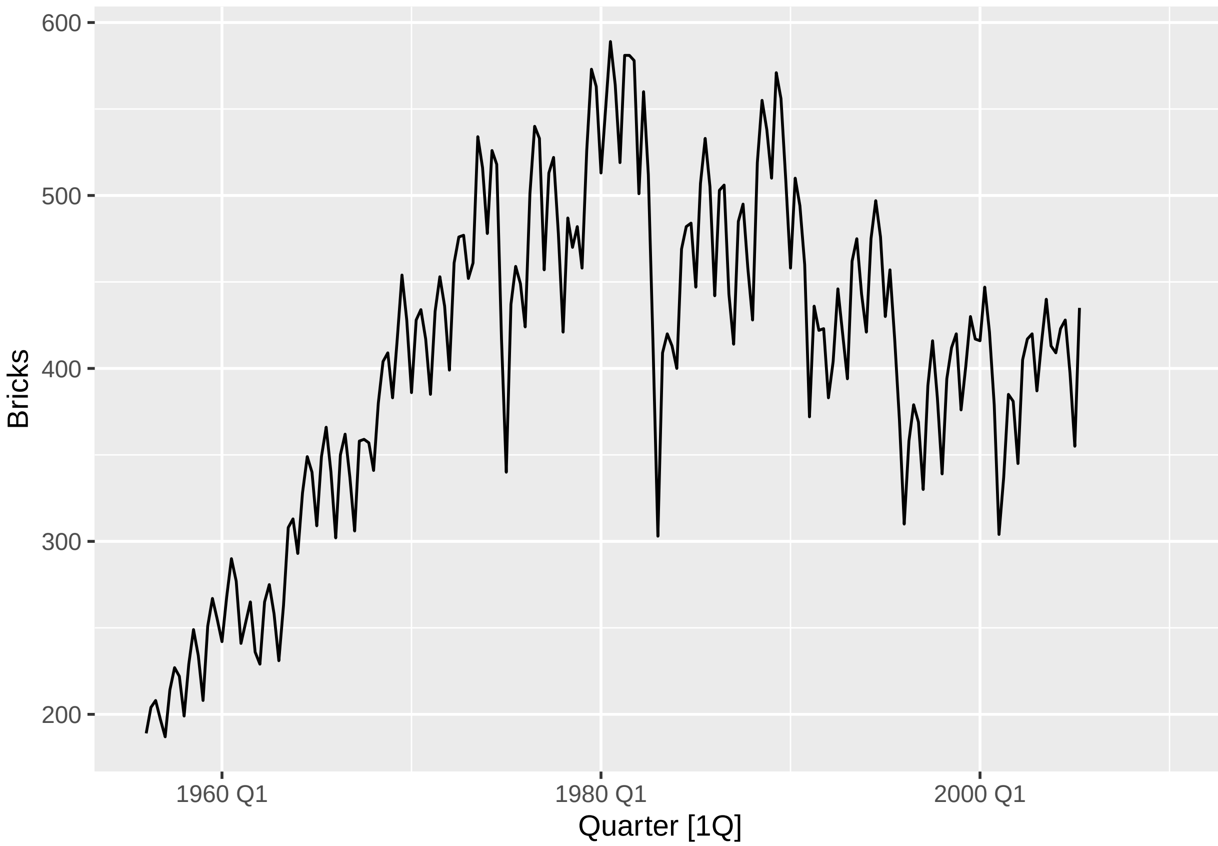
- Use ? (or help()) to find out about the data in each series.
- What is the time interval of each series?
- Use autoplot() to produce a time plot of each series.
- For the last plot, modify the axis labels and title.

Bricks* from *aus_production

The time interval in the *aus_production* dataset is quarterly.

```
#help(aus_production)
autoplot(aus_production, Bricks)
```

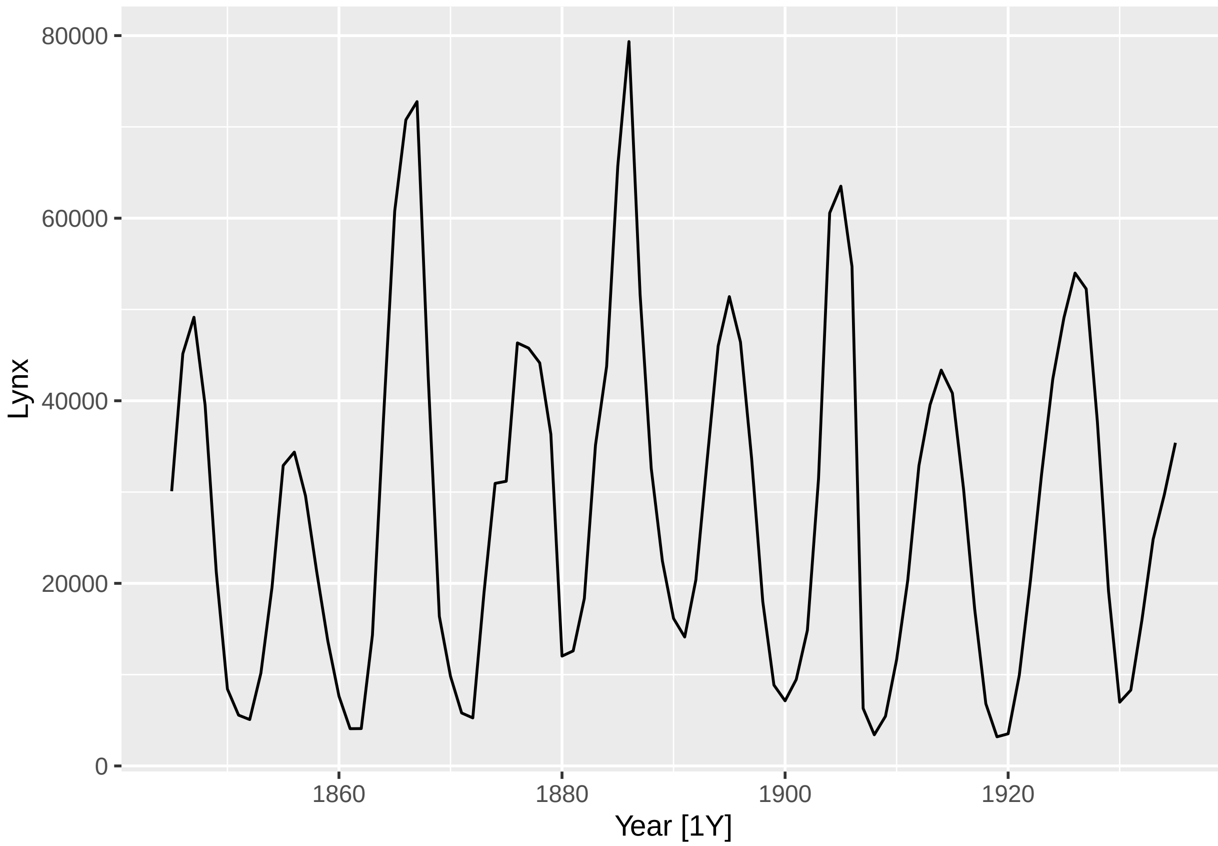
```
## Warning: Removed 20 rows containing missing values (`geom_line()`).
```



Lynx* from *pelt

The time interval for the *pelt* dataset is annual.

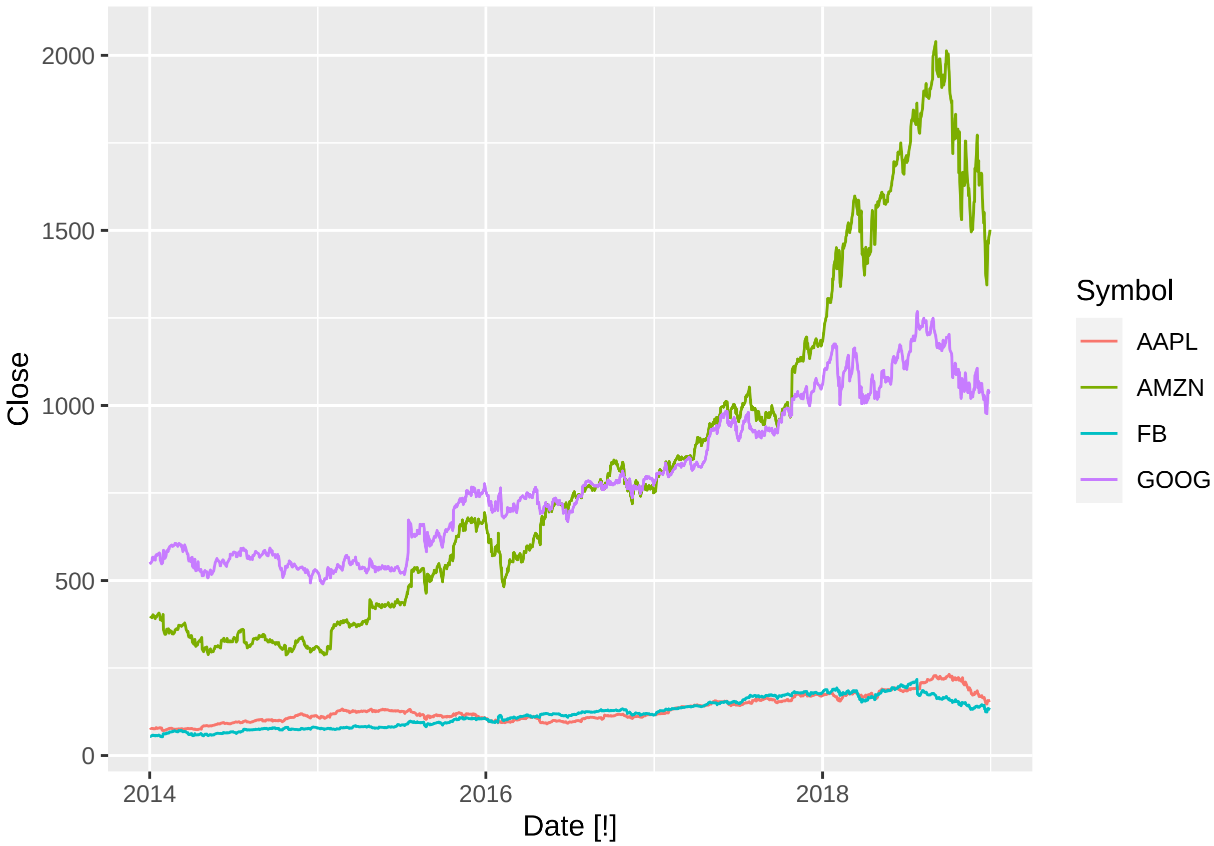
```
#help(pelt)
autoplot(pelt, Lynx)
```



Close* from *gafa_stock

The time interval for the *gafa_stock* dataset is trading days of the stock market.

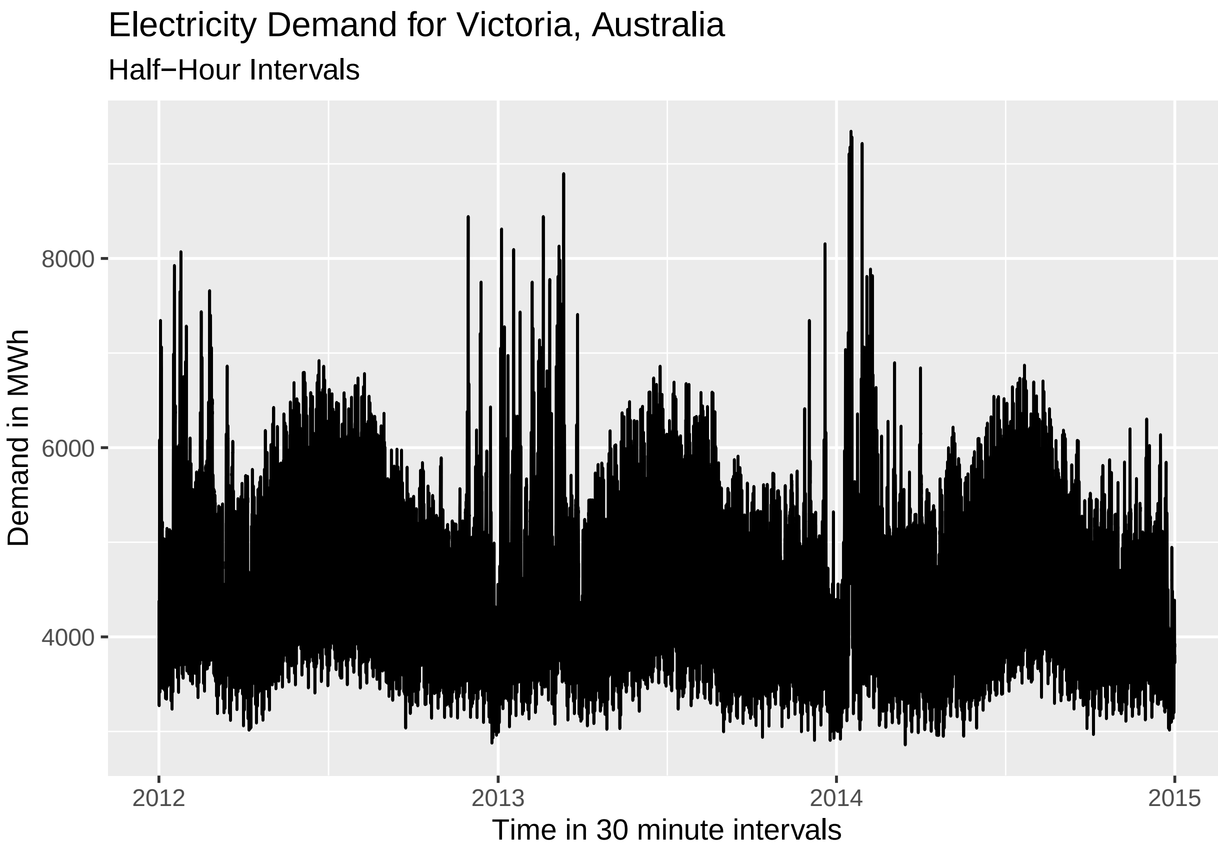
```
#help(gafa_stock)
autoplot(gafa_stock, Close)
```



Demand* from *vic_elec

The time interval for the *vic_elec* dataset is every 30 minutes.

```
help(vic_elec)
autoplot(vic_elec, Demand) +
  ggtitle('Electricity Demand for Victoria, Australia',
          subtitle = 'Half-Hour Intervals') +
  labs(x = 'Time in 30 minute intervals',
       y = 'Demand in MWh')
```



Exercise 2.2

Use *filter()* to find what days correspond to the peak closing price for each of the four stocks in *gafa_stock*.

```
gafa_stock %>%
  group_by(Symbol) %>%
  filter(Close == max(Close)) %>%
  select(Symbol,Date,Close)
```

```
## # A tsibble: 4 x 3 [!]
## # Key:       Symbol [4]
## # Groups:    Symbol [4]
##   Symbol Date       Close
##   <chr>  <date>     <dbl>
## 1 AAPL   2018-10-03  232.
## 2 AMZN   2018-09-04 2040.
## 3 FB     2018-07-25  218.
## 4 GOOG   2018-07-26 1268.
```

Exercise 2.3

Download the file tute1.csv from the book website, open it in Excel, 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.

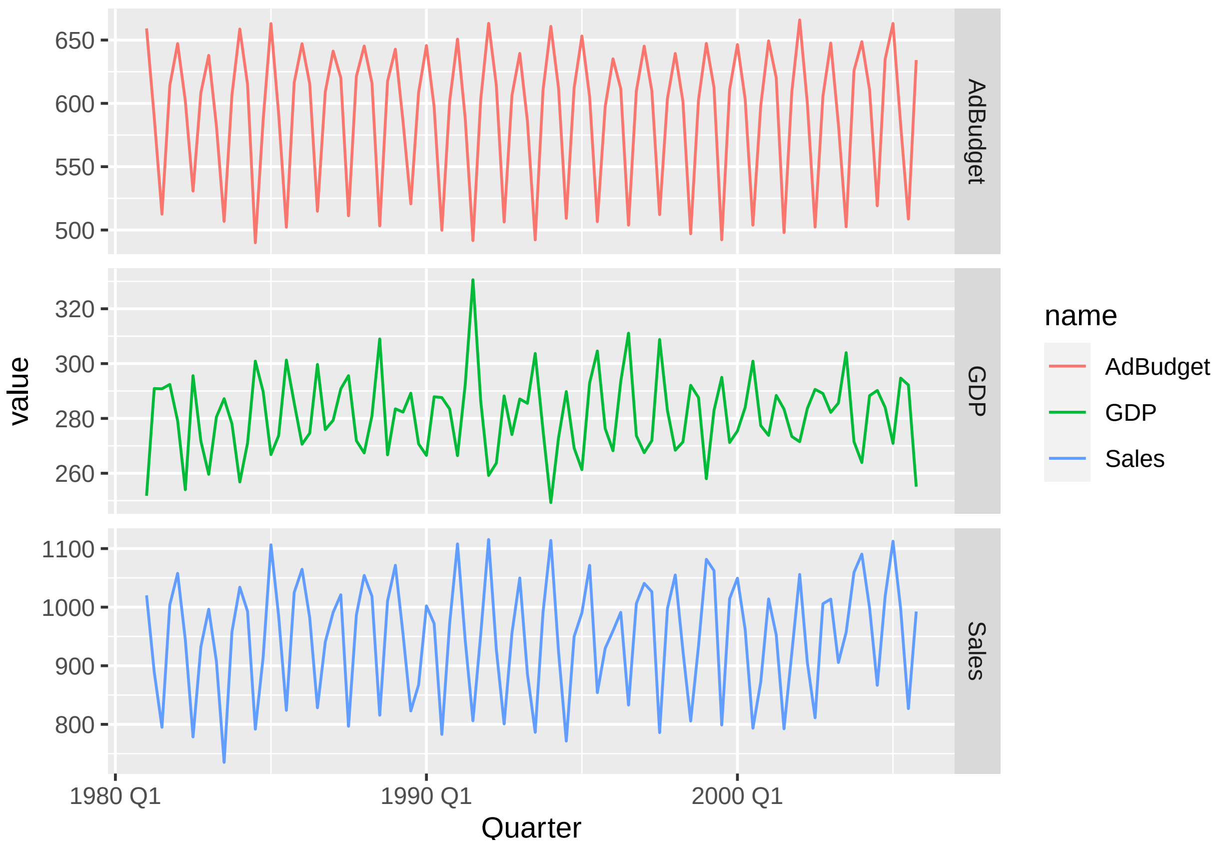
Downloaded the tute1.csv file and opened it.

Read tute1.csv from github folder and convert to a time series.

```
ts_tute <- read.csv('https://raw.githubusercontent.com/dab31415/DATA624/main/Homework/tute1.csv') %>%
  mutate(Quarter = yearquarter(Quarter)) %>%
  as_tsibble(index = Quarter)
```

Construct plot.

```
ts_tute %>%
  pivot_longer(-Quarter) %>%
  ggplot(aes(x = Quarter, y = value, colour = name)) +
  geom_line() +
  facet_grid(name ~ ., scales = 'free_y')
```



The facet_grid function separtes the graph into individual panels based on the variable provided.

Exercise 2.4

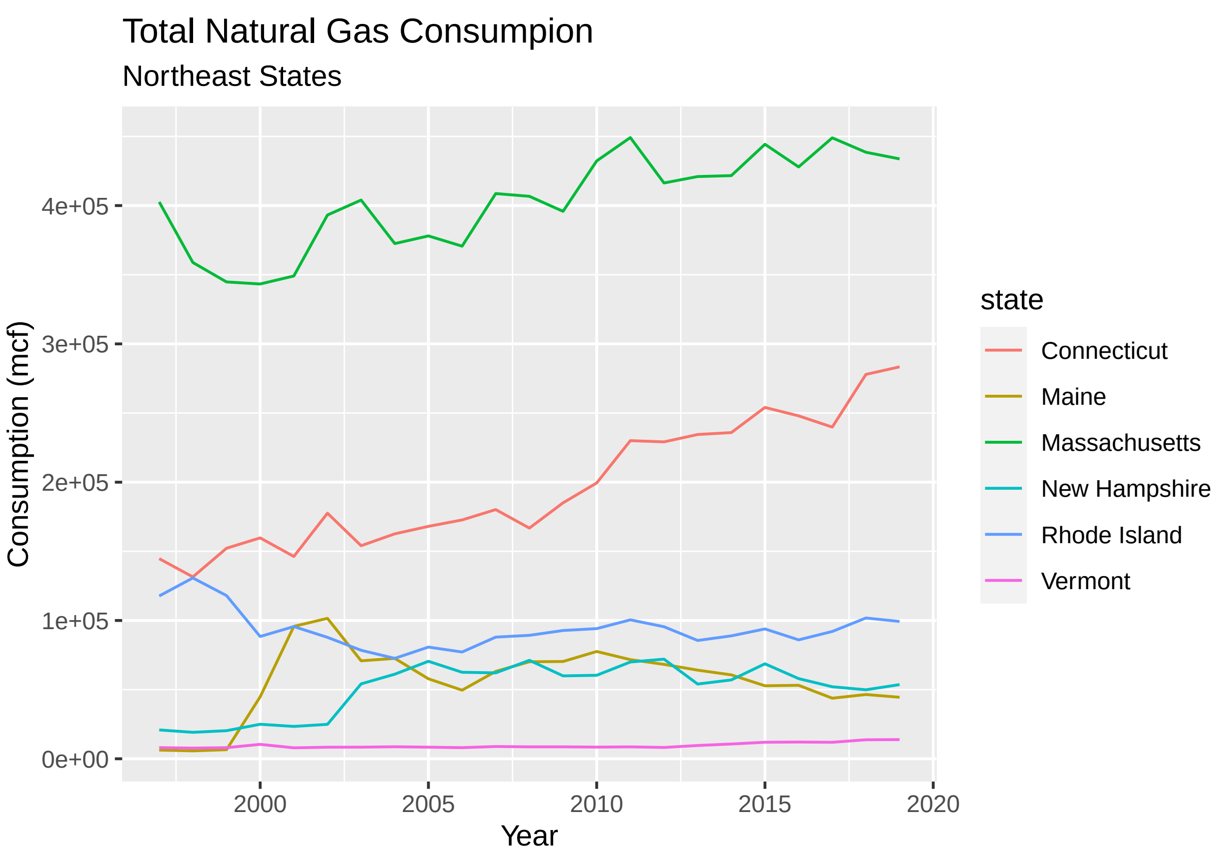
The USgas package contains data on the demand for natural gas in the US.

a. Install the USgas package.

b. Create a *tsibble* from *us_total* with year as the index and state as the key.

c. Plot the annual natural gas consumption by state for the New England area (comprising the states of Maine, Vermont, New Hampshire, Massachusetts, Connecticut and Rhode Island).

```r
library(USgas)

ts_usgas <- us_total %>%
  as_tsibble(index = year, key = state)

ne_states <- c('Maine','Vermont','New Hampshire', 'Massachusetts', 'Connecticut', 'Rhode Island')

ts_usgas %>%
  filter(state %in% ne_states) %>%
  autoplot(y) +
    labs(title = 'Total Natural Gas Consumpion',
         subtitle = 'Northeast States',
         x = 'Year',
         y = 'Consumption (mcf)')
```



Exercise 2.5

a. Download tourism.xlsx from the book website and read it into R using *readxl::read_excel()*.
b. Create a tsibble which is identical to the *tourism* tsibble from the *tsibble* package.

```r
library(readxl)

xls_tourism <- read_excel('./tourism.xlsx')
ts_tourism <- xls_tourism %>%
  mutate(Quarter = yearquarter(Quarter)) %>%
  as_tsibble(index = Quarter, key = c('Region','Purpose','State'))

head(ts_tourism)
```

```
## # A tsibble: 6 x 5 [1Q]
## # Key:       Region, State, Purpose [1]
##   Quarter Region   State           Purpose  Trips
##     <qtr> <chr>    <chr>           <chr>    <dbl>
## 1 1998 Q1 Adelaide South Australia Business  135.
## 2 1998 Q2 Adelaide South Australia Business  110.
## 3 1998 Q3 Adelaide South Australia Business  166.
## 4 1998 Q4 Adelaide South Australia Business  127.
## 5 1999 Q1 Adelaide South Australia Business  137.
## 6 1999 Q2 Adelaide South Australia Business  200.
```

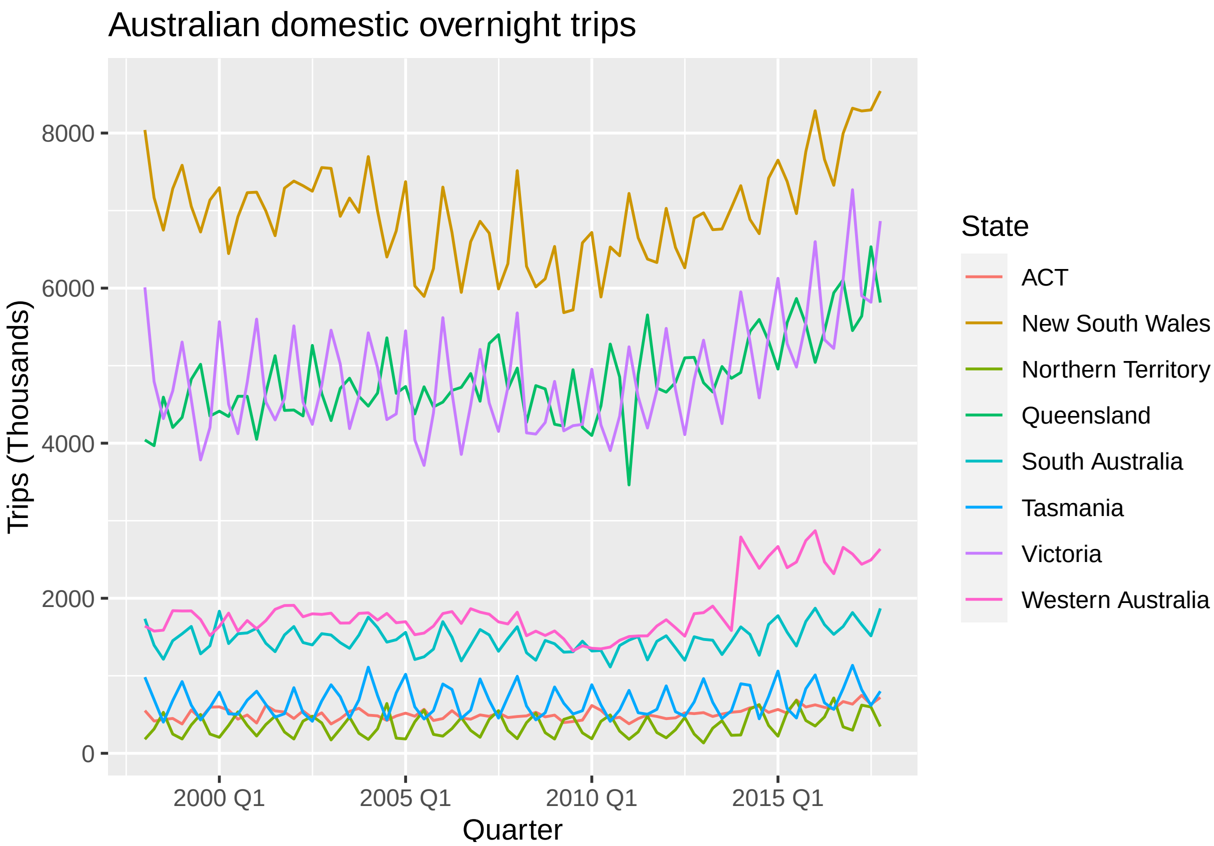
c. Find what combination of *Region* and *Purpose* had the maximum number of overnight trips on average.

```r
ts_tourism %>%
  mutate(mean_trips = mean(Trips),
         .by = c(Region,Purpose), .keep = 'none') %>%
  distinct() %>%
  top_n(1, mean_trips)
```

```
## # A tibble: 1 x 3
##   Region Purpose  mean_trips
##   <chr>  <chr>         <dbl>
## 1 Sydney Visiting       747.
```

d. Create a new tsibble which combines the Purposes and Regions, and just has total trips by State.

```r
ts_state_tourism <- xls_tourism %>%
  mutate(state_trips = sum(Trips),
         Quarter = yearquarter(Quarter),
         .by = c(Quarter, State), .keep = 'none') %>%
  distinct() %>%
  as_tsibble(index = Quarter, key = State)

ts_state_tourism %>%
  autoplot(.vars = state_trips) +
    labs(title = 'Australian domestic overnight trips',
         x = 'Quarter',
         y = 'Trips (Thousands)')
```



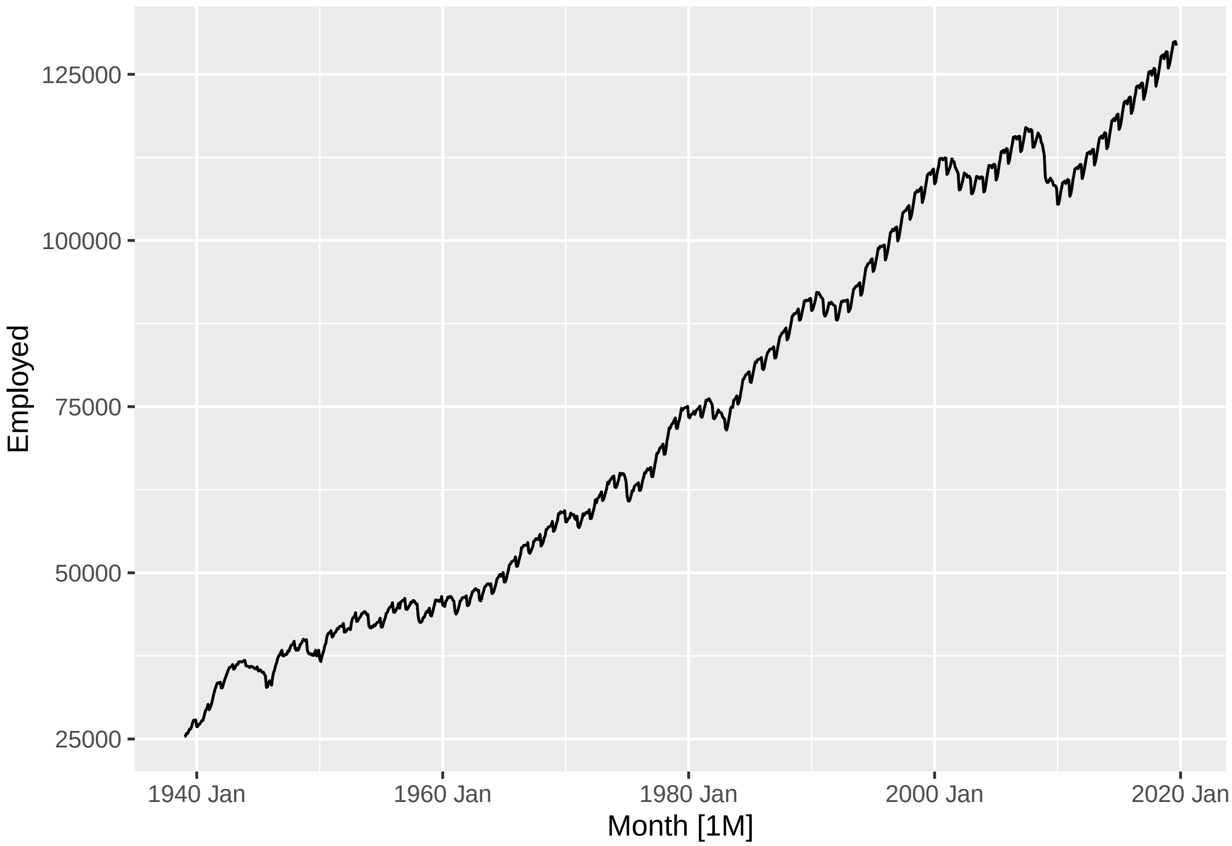
Exercise 2.8

Use the following graphics functions: autoplot(), gg_season(), gg_subseries(), gg_lag(), ACF() and explore features from the following time series: "Total Private" *Employed* from *us_employment*, *Bricks* from *aus_production*, *Hare* from *pelt*, "H02" *Cost* from *PBS*, and *Barrels* from *us_gasoline*.

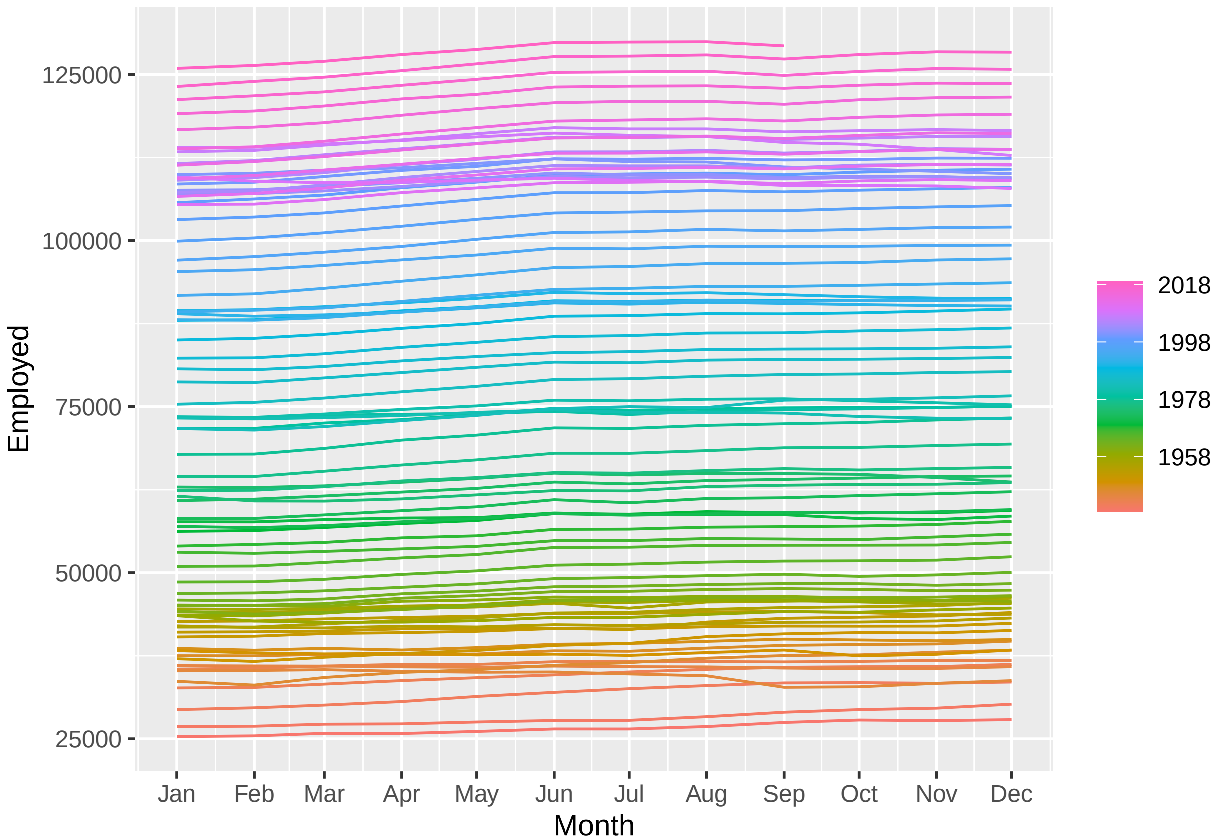
- Can you spot any seasonality, cyclicity, and trend?
- What do you learn about the series?
- What can you say about the seasonal patterns?
- Can you identify any unusual years?

"Total Private" *Employed* from *us_employment*

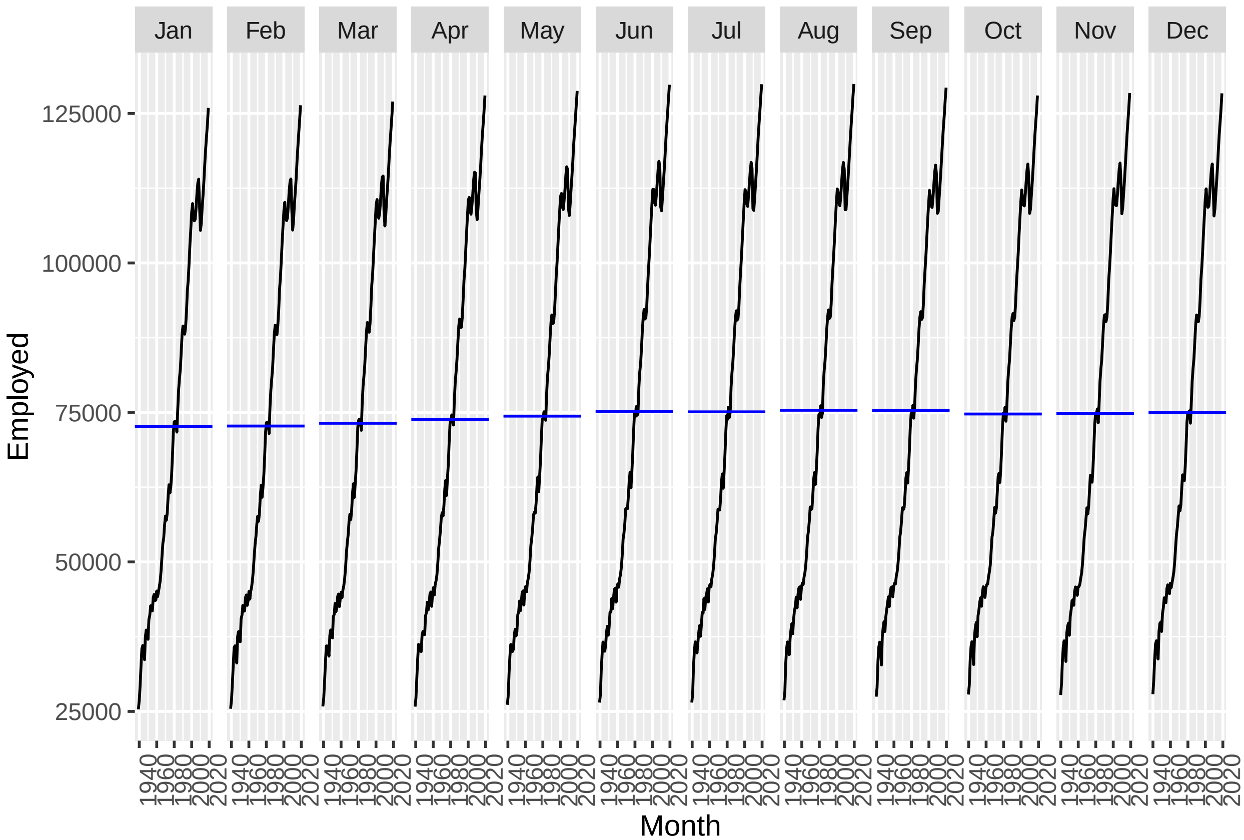
```
ts_employed <- us_employment %>%
  filter(Title == 'Total Private') %>%
  select(Month, Employed)

autoplot(ts_employed, .vars = Employed)
```



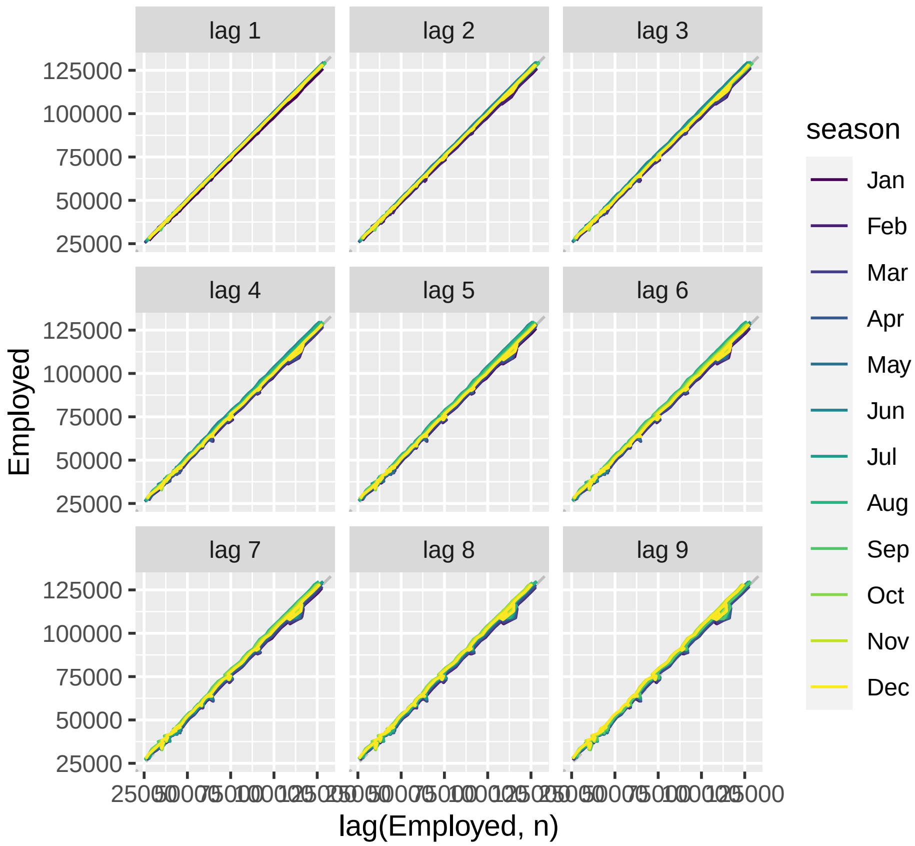

```
gg_season(ts_employed, y = Employed)
```



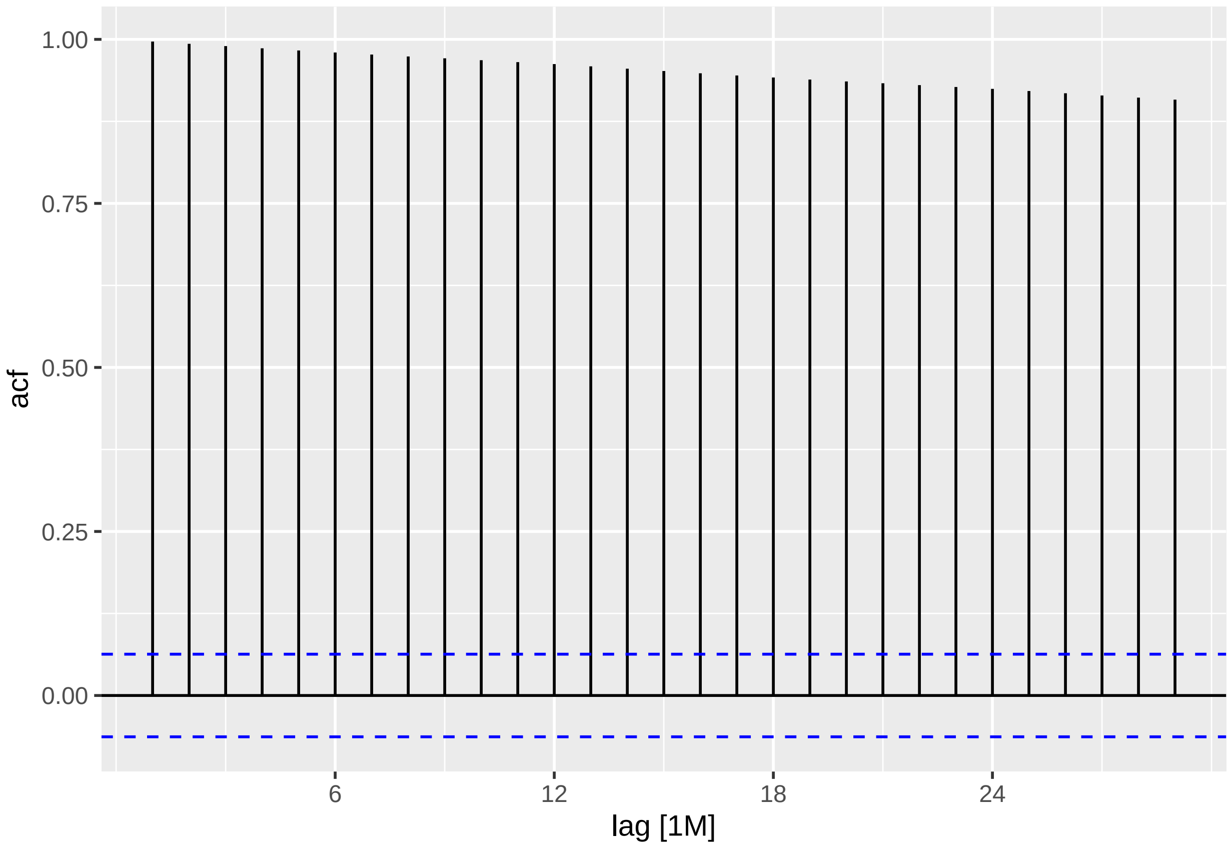

```r
gg_subseries(ts_employed, y = Employed)
```




```
gg_lag(ts_employed, y = Employed)
```




```
ACF(ts_employed, y = Employed) %>%
  autoplot()
```

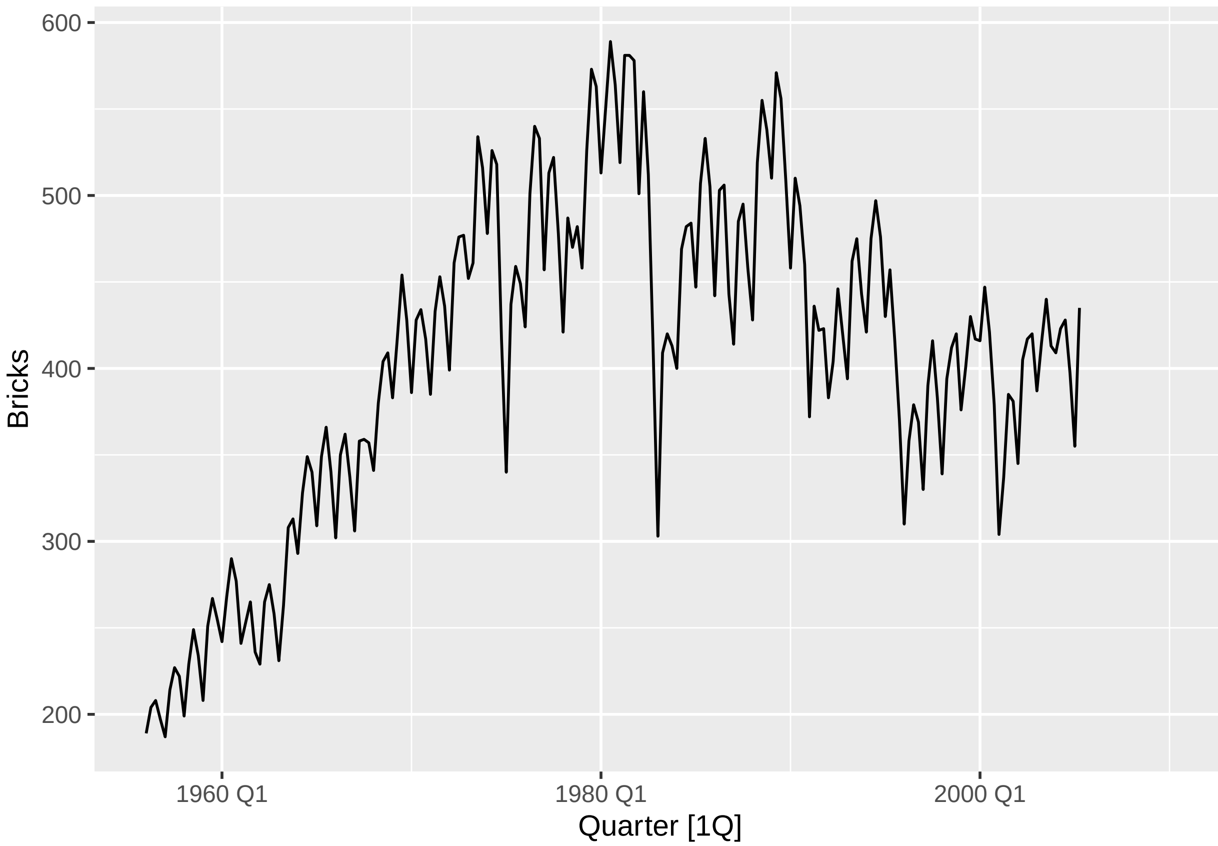


The graphs of employment data show an annual seasonality with peaks in the summer months. There is an overall trend that is increasing over time. There are a few years with decreases in employment which are against the long term trend, with 2008 having the most significant decrease.

Bricks* from *aus_production

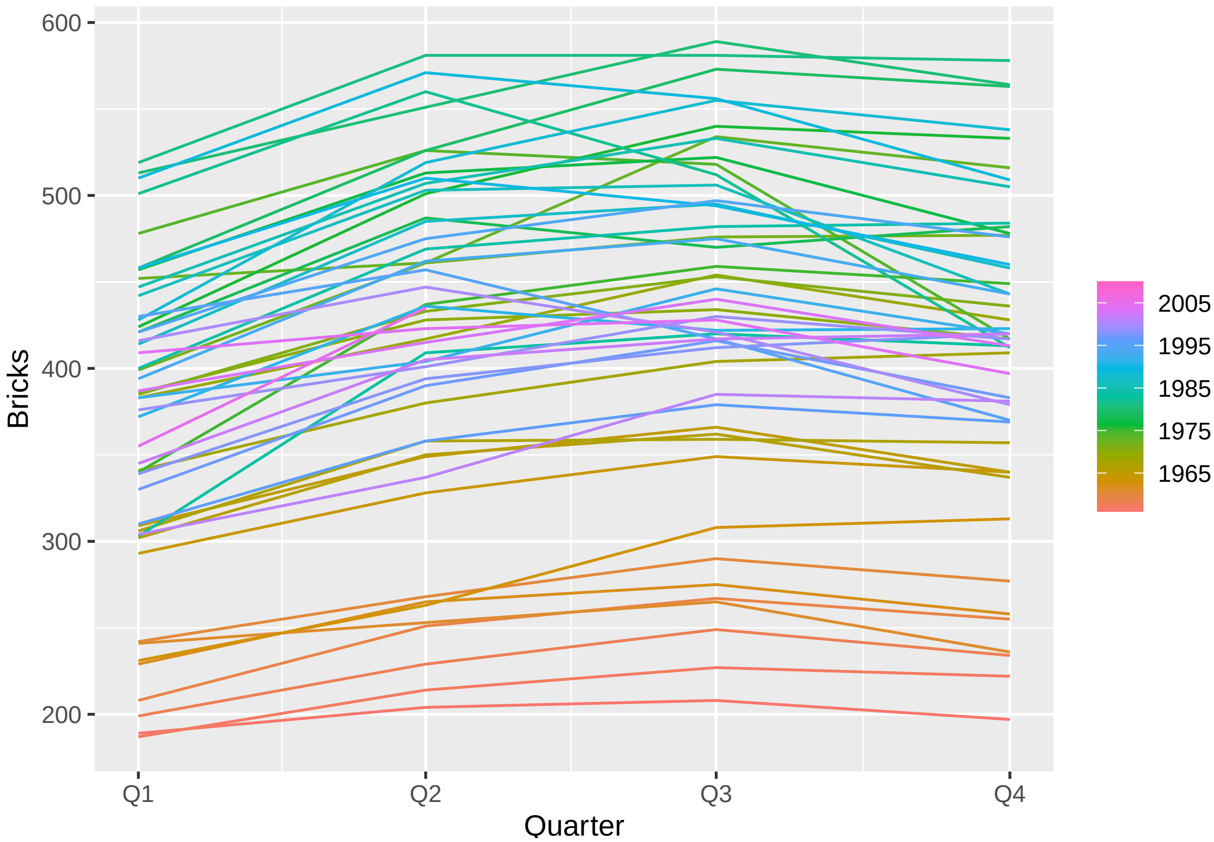
```
ts_bricks <- aus_production %>%
  select(Quarter, Bricks)

autoplot(ts_bricks, .vars = Bricks)
```

```
## Warning: Removed 20 rows containing missing values (‘geom_line()‘).
```



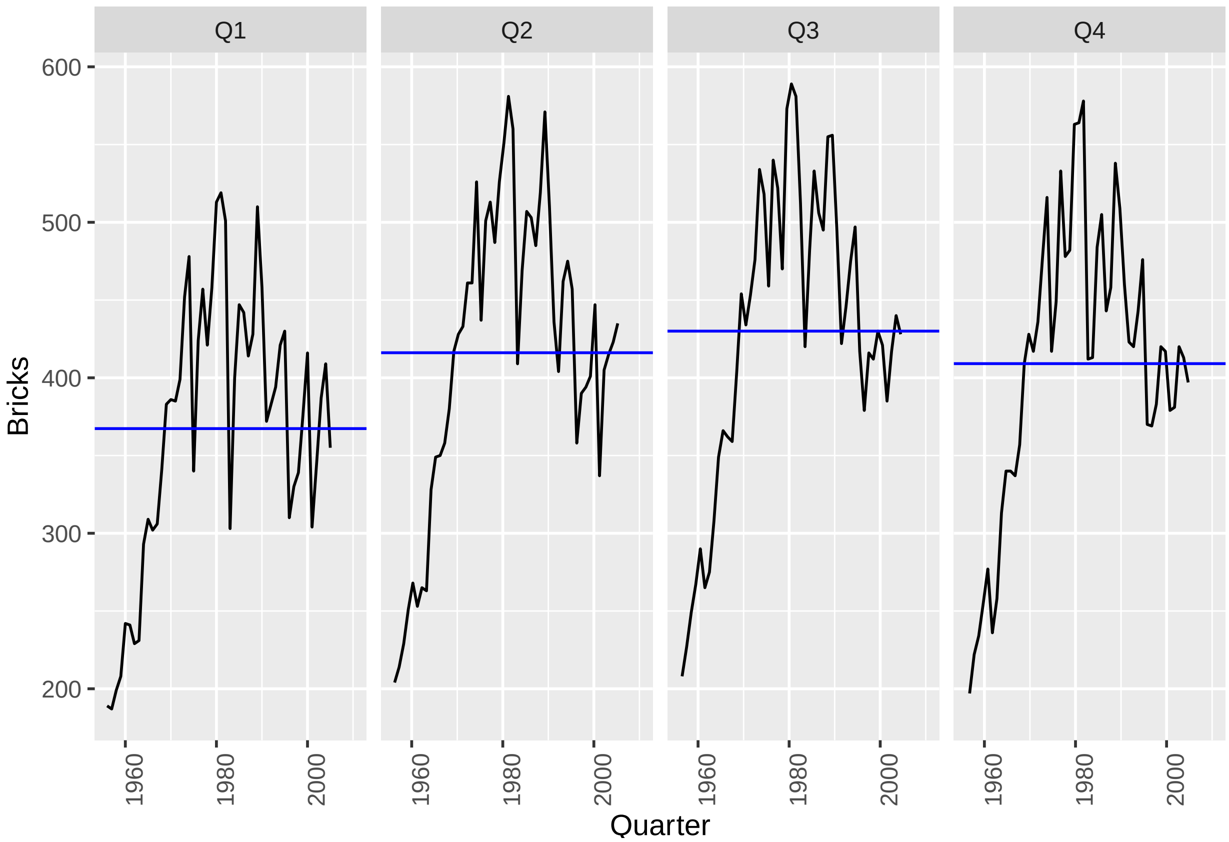

```
gg_season(ts_bricks, y = Bricks)
```

```
## Warning: Removed 20 rows containing missing values (`geom_line()`).
```



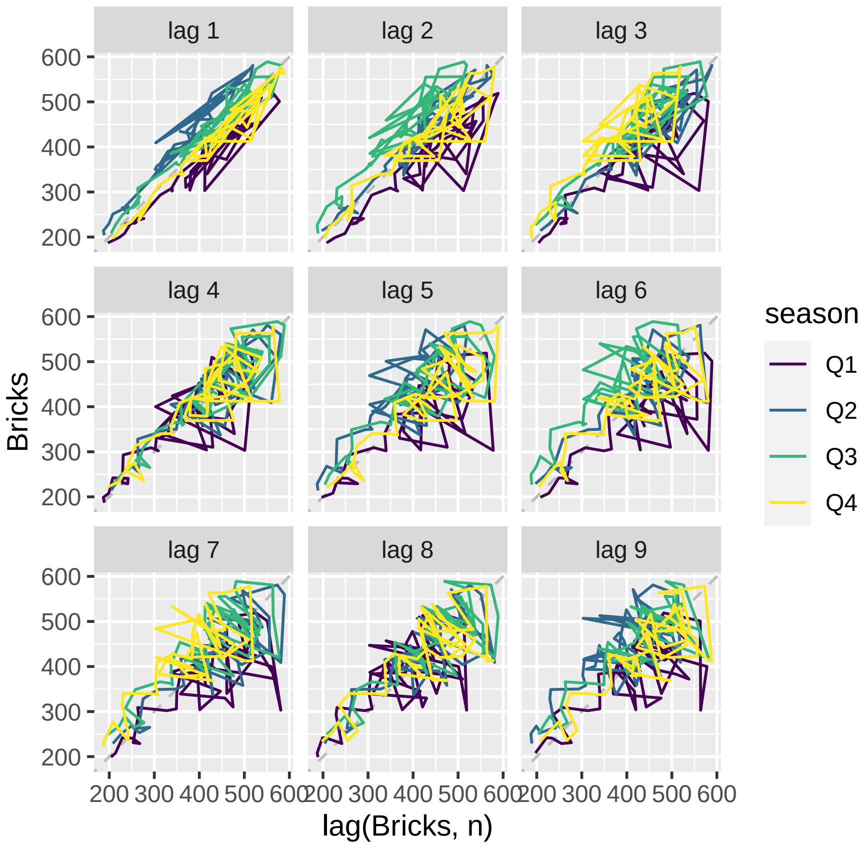

```r
gg_subseries(ts_bricks, y = Bricks)
```

```
## Warning: Removed 5 rows containing missing values (`geom_line()`).
```

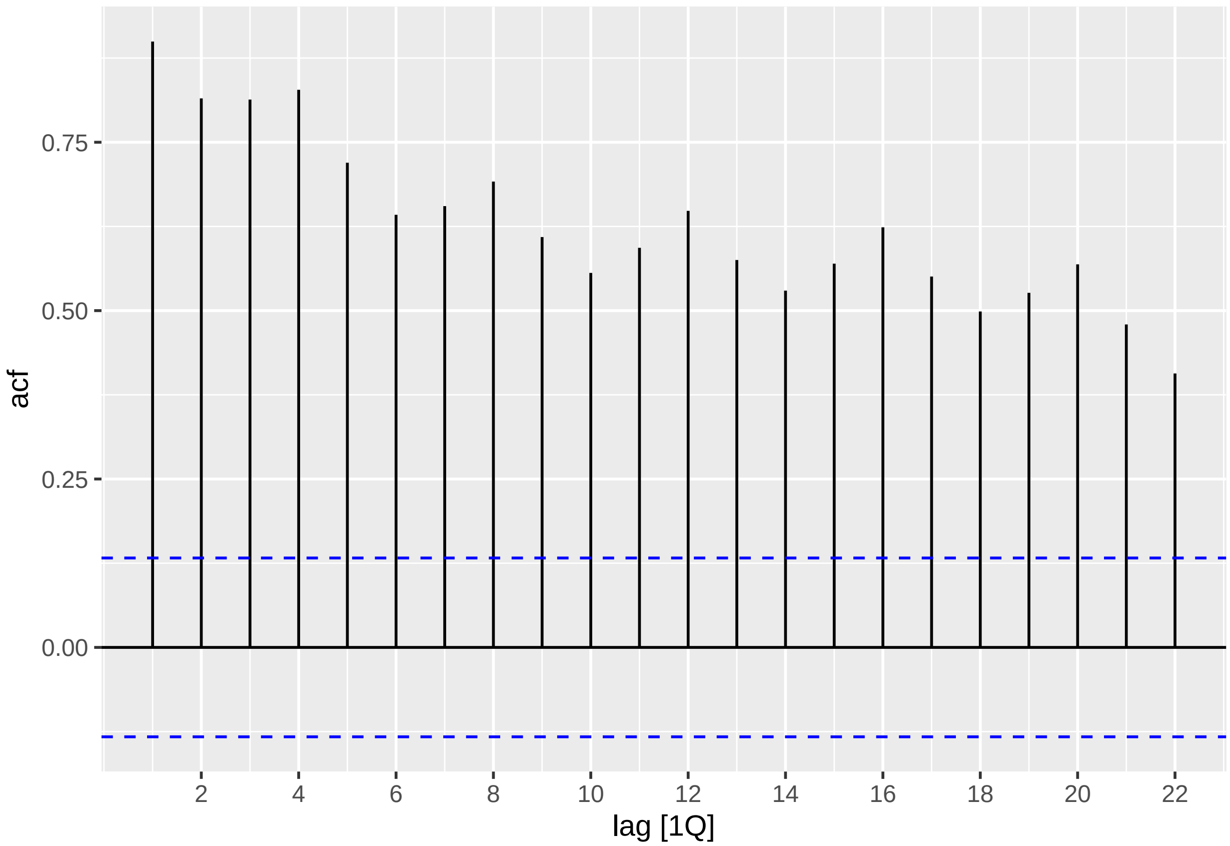



```
gg_lag(ts_bricks, y = Bricks)
```

```
## Warning: Removed 20 rows containing missing values (gg_lag).
```



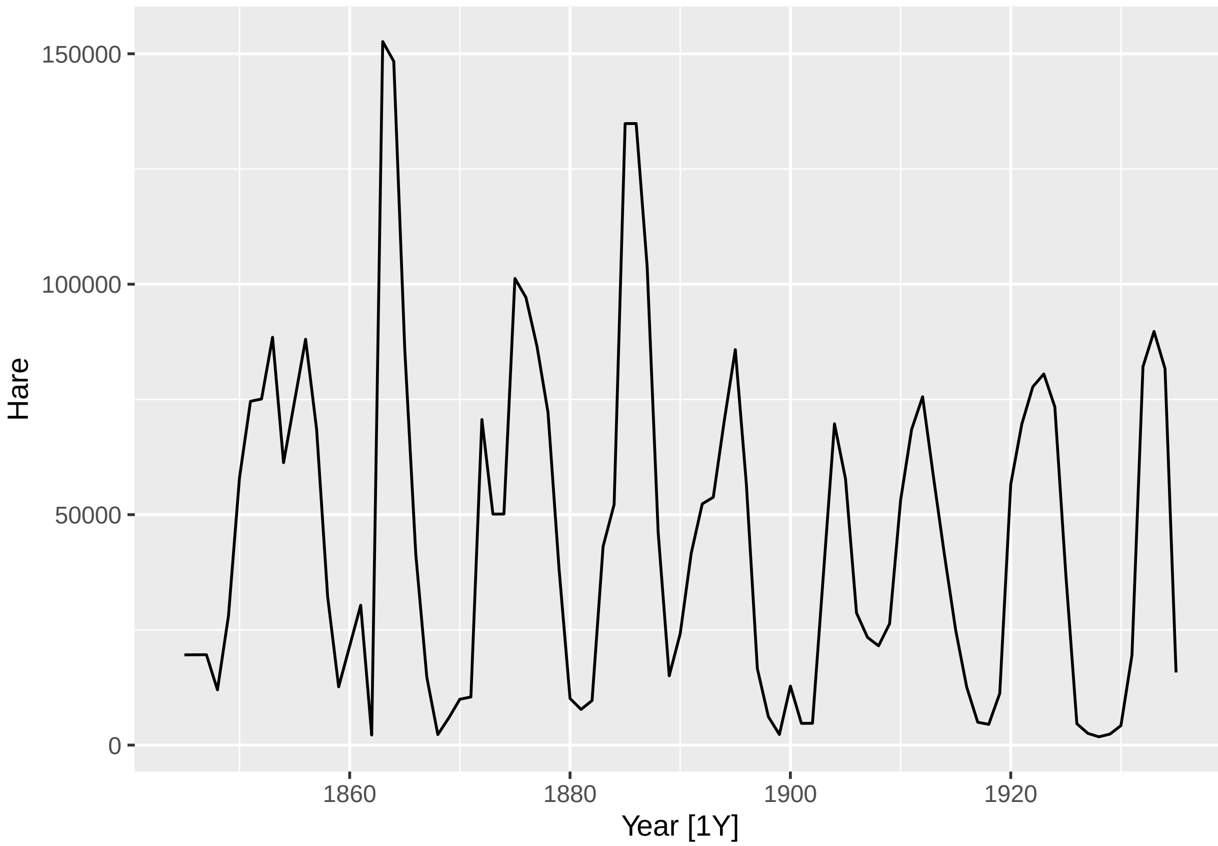

```
ACF(ts_bricks, y = Bricks) %>%
  autoplot()
```



The overall trend was increasing until about 1982, since that time the trend has been decreasing. The data shows a seasonality which repeats annually. Brick production is lowest in Q1 summer in Australia.

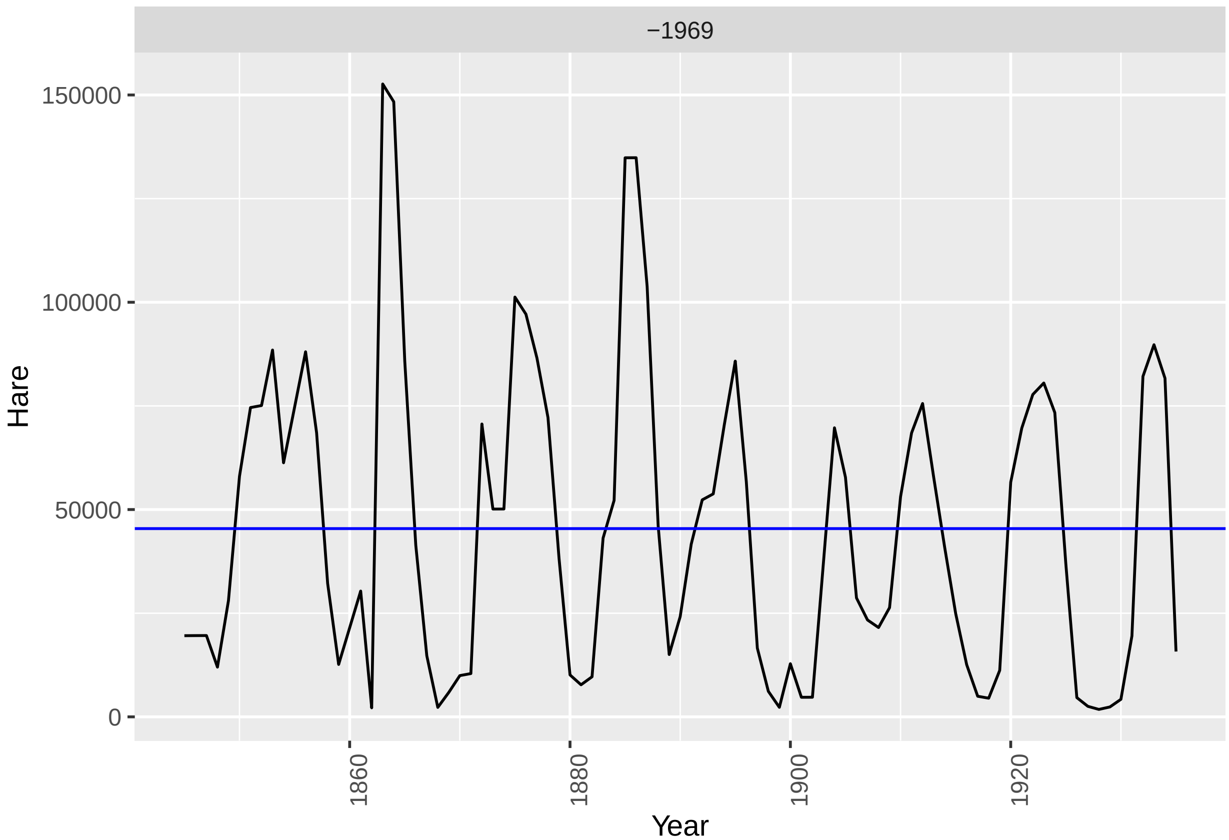
Hare* from *pelt

```
ts_hare <- pelt %>%
  select(Year, Hare)

autoplot(ts_hare, .vars = Hare)
```

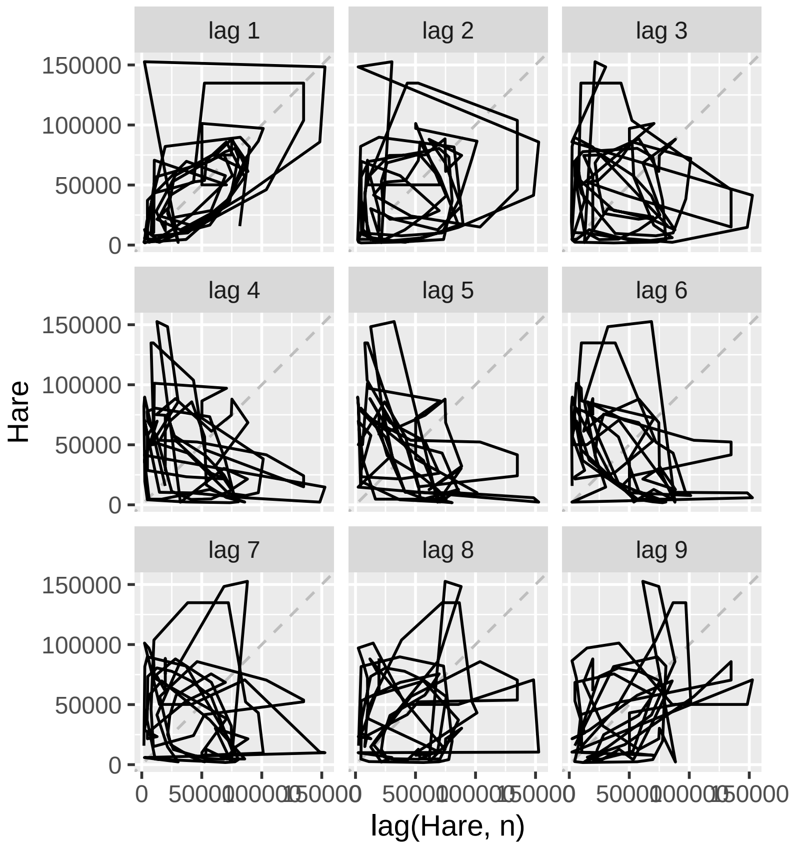



```r
#gg_season(ts_hare, y = Hare)
```

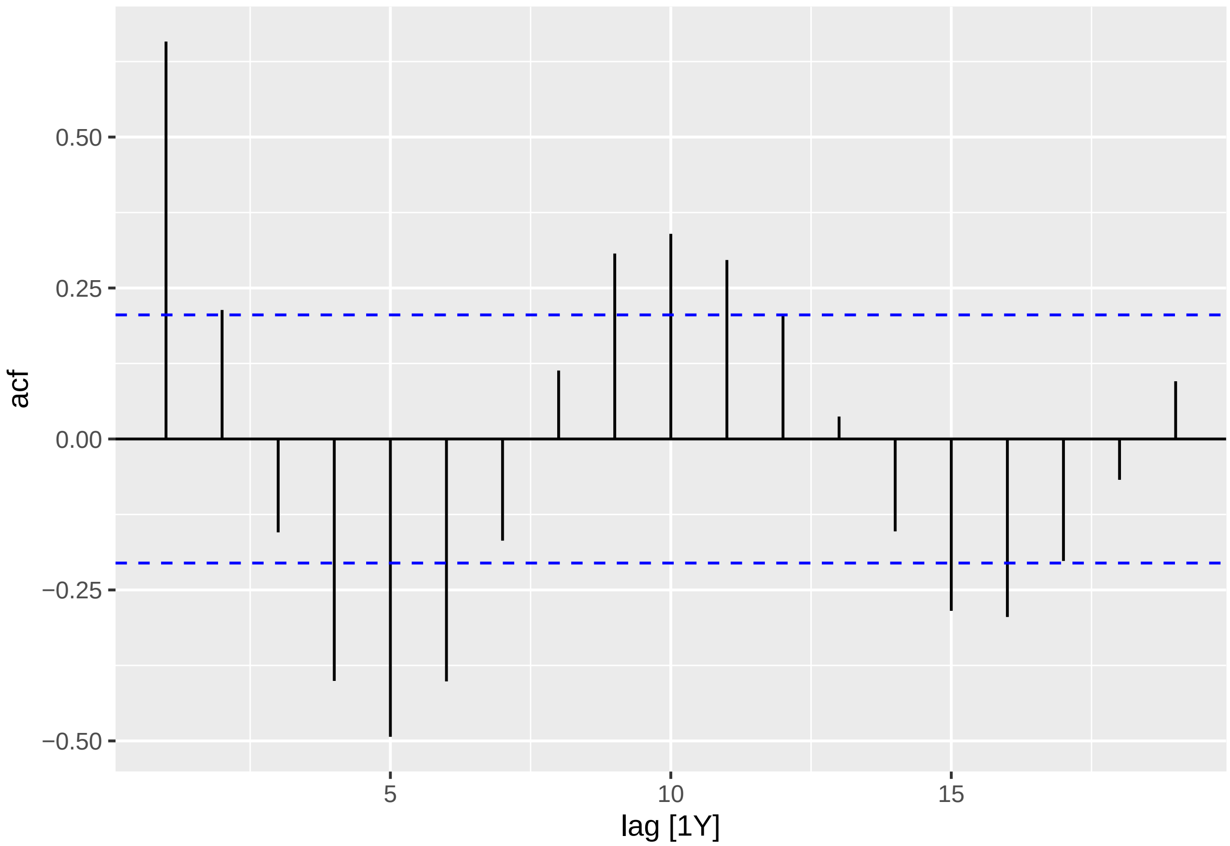
```r
gg_subseries(ts_hare, y = Hare)
```




```r
gg_lag(ts_hare, y = Hare)
```



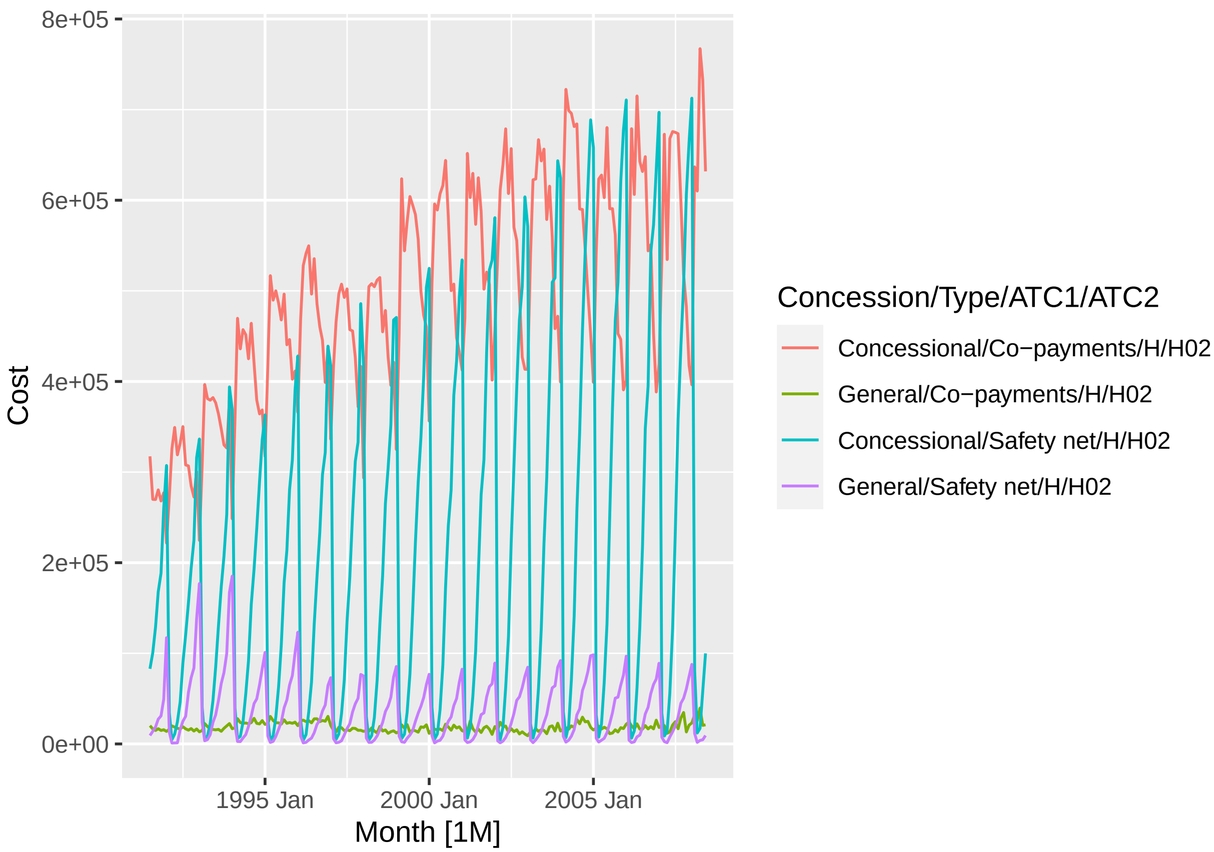

```r
ACF(ts_hare, y = Hare) %>%
  autoplot()
```



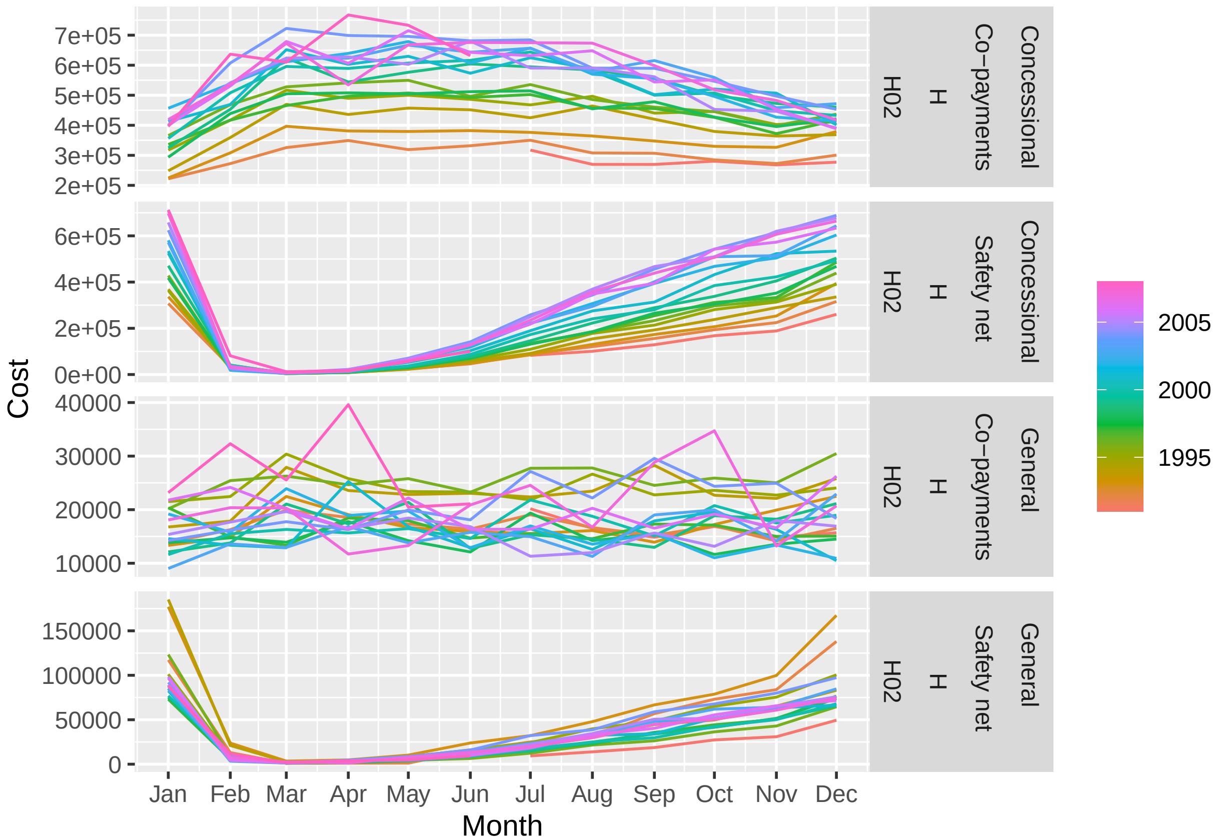
The seasonality of Snowshoe Hare pelts is about 8 years. The overall trend of the data is flat.

"H02" ***Cost*** **from** ***PBS***

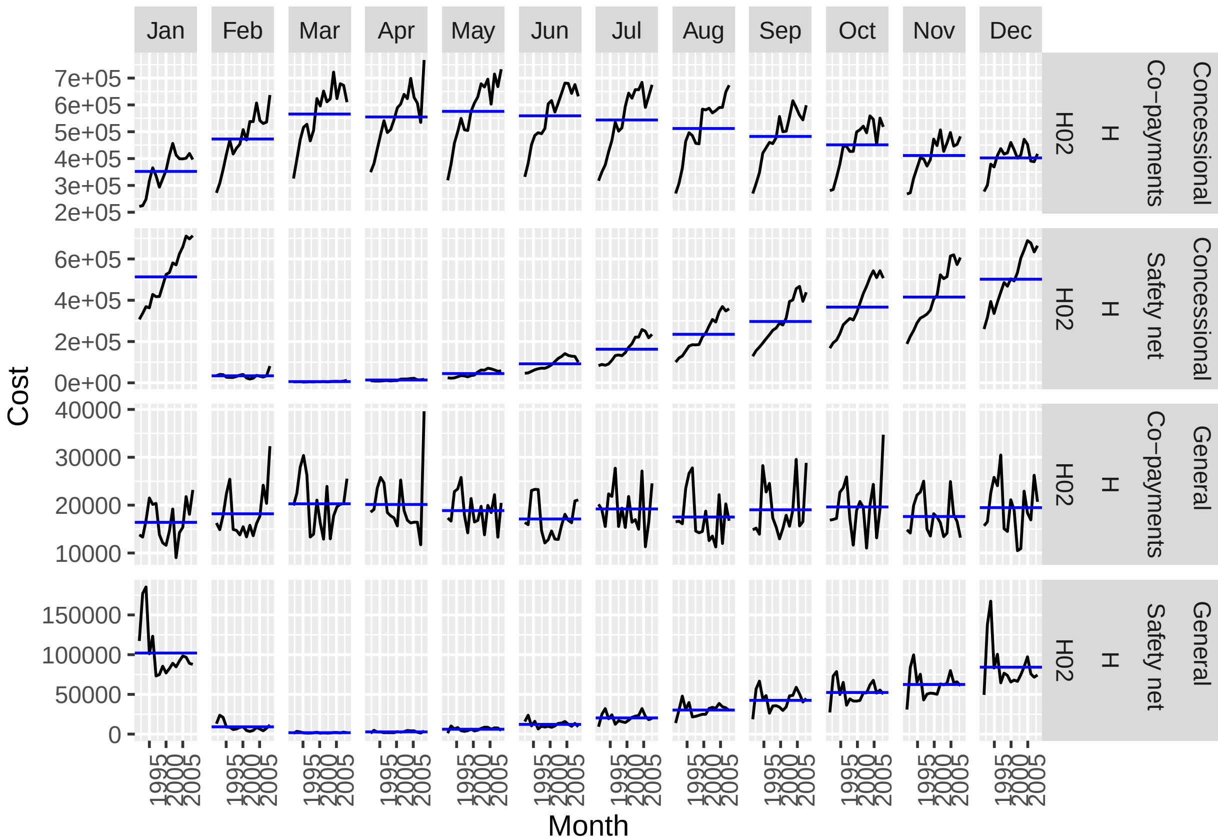
```
ts_cost <- PBS %>%
  filter(ATC2 == 'H02') %>%
  select(Month, Cost)

autoplot(ts_cost, .vars = Cost)
```



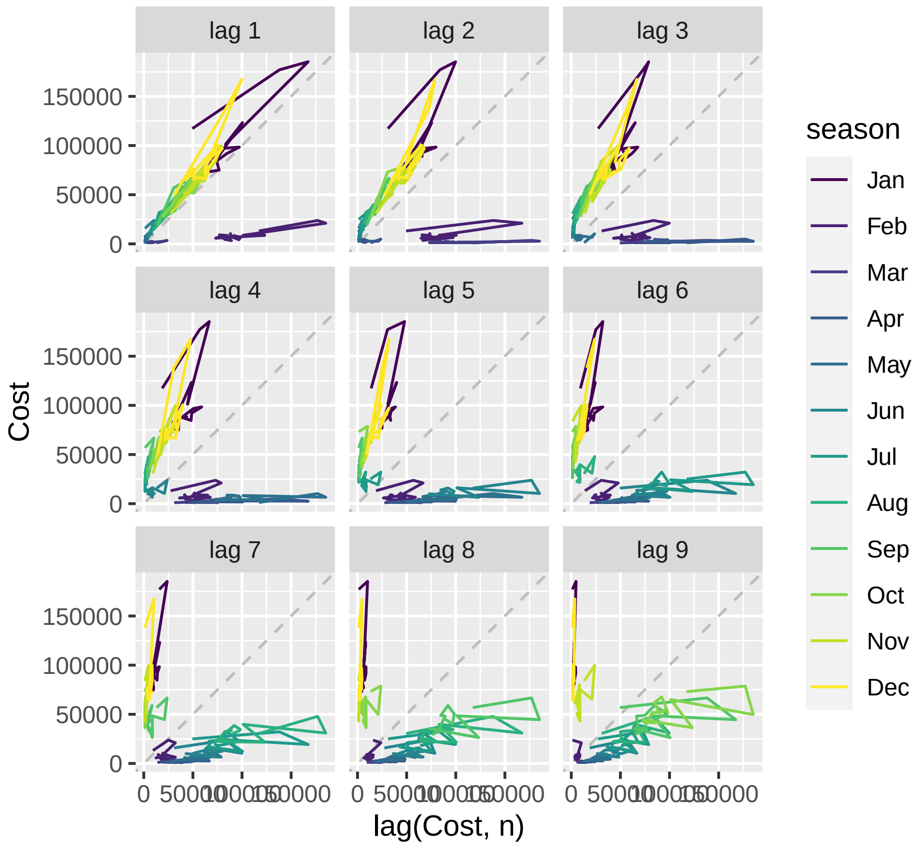

```
gg_season(ts_cost, y = Cost)
```



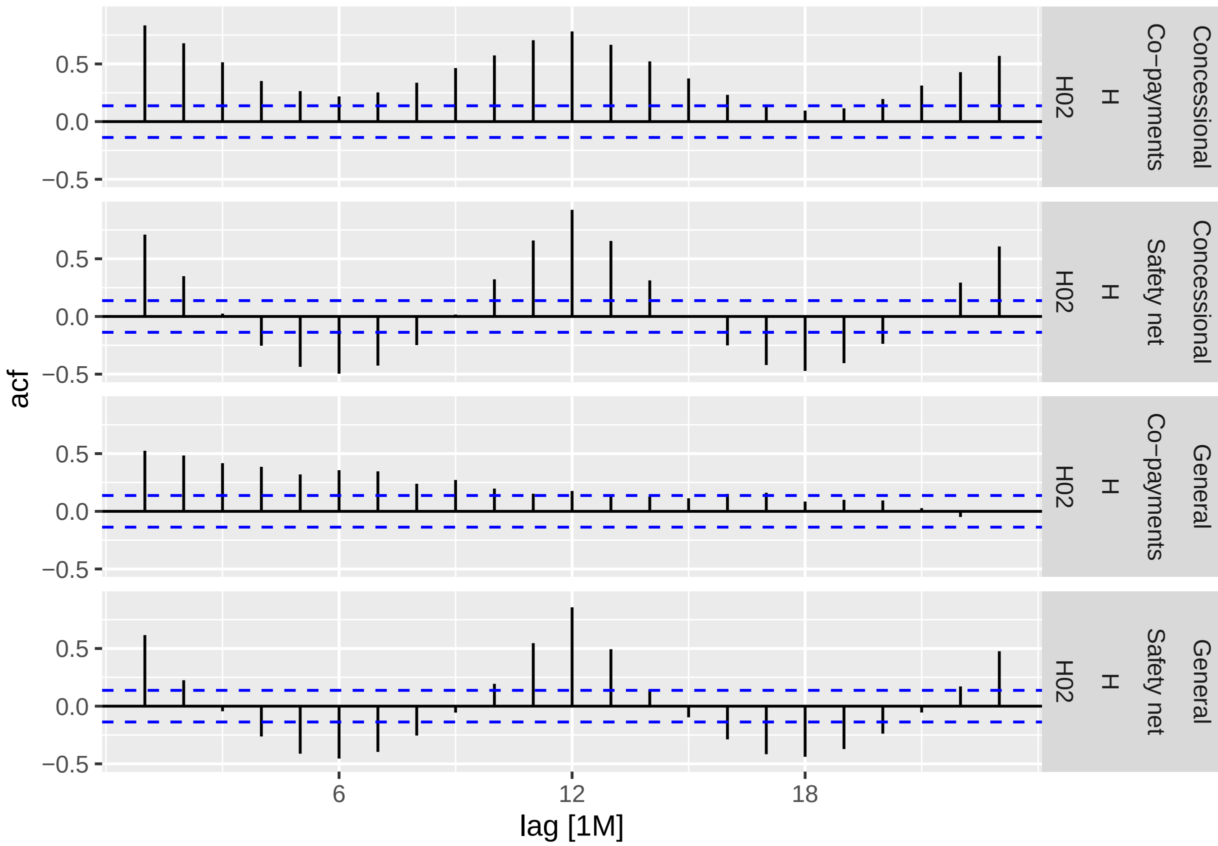

```r
gg_subseries(ts_cost, y = Cost)
```




```
gg_lag(ts_cost %>% filter(Concession == 'General', Type == 'Safety net'), y = Cost)
```



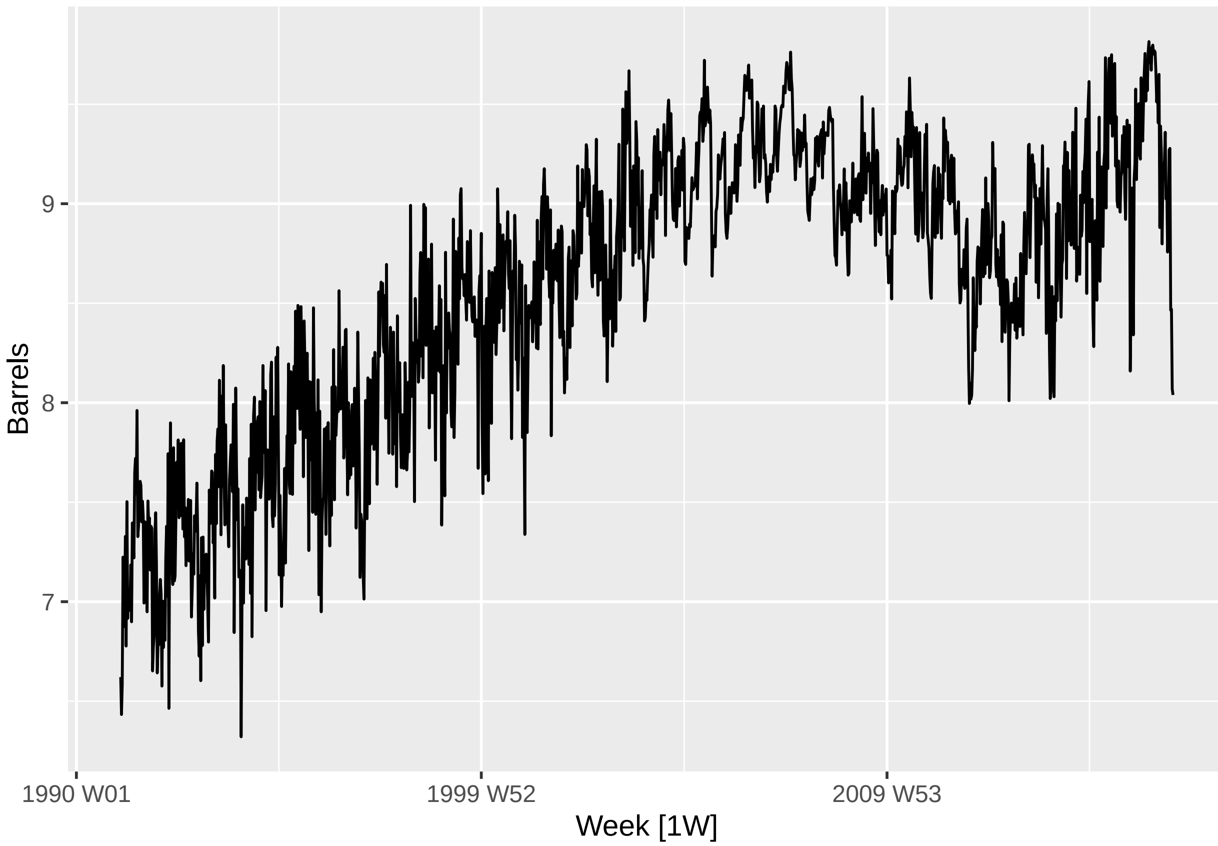

```
ACF(ts_cost, y = Cost) %>%
  autoplot()
```



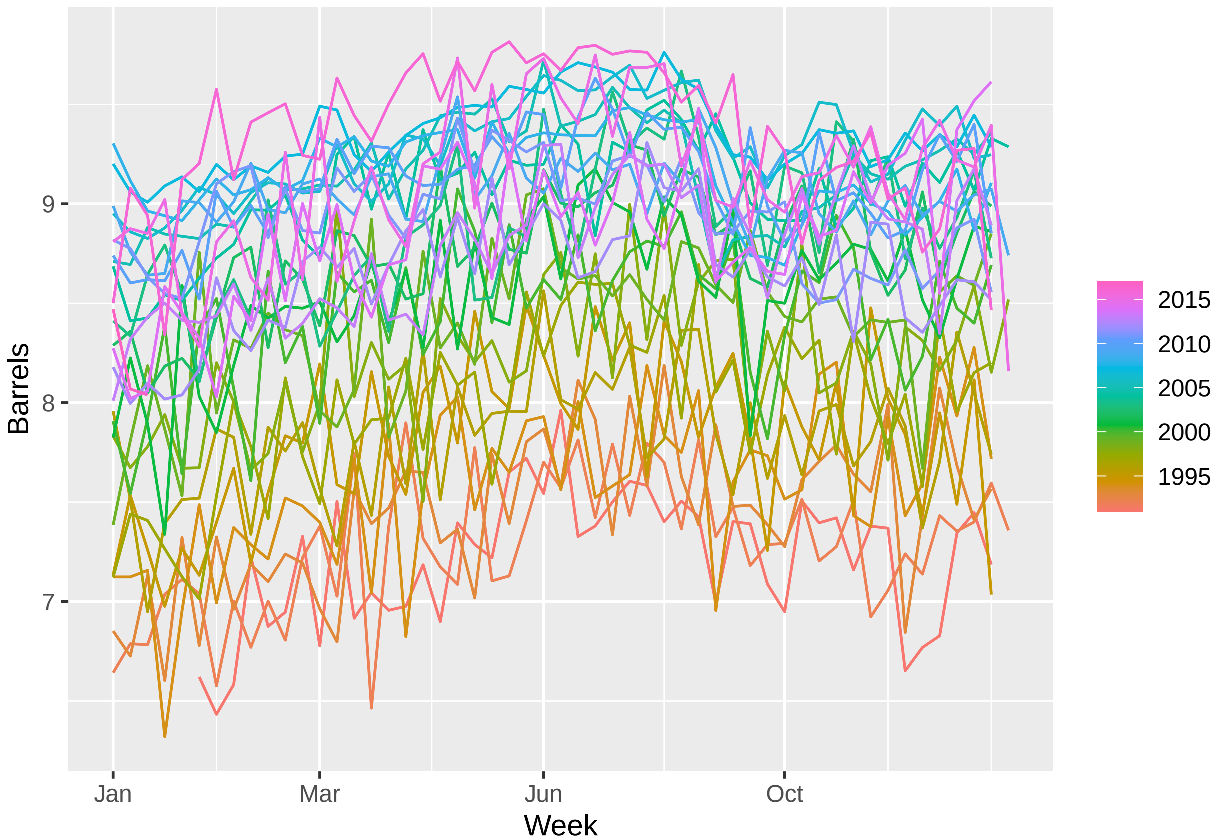
There annual seasonality of the data series due to the reset of annual deductibles. At the beginning of the year, co-payments are paid, but the safety net funds remain low until the patient deductibles are met. There is a bit of lag in the data as January payment are likely the result of December prescriptions. There is an overall trend in the data is increasing.

Barrels **from** ***us_gasoline***

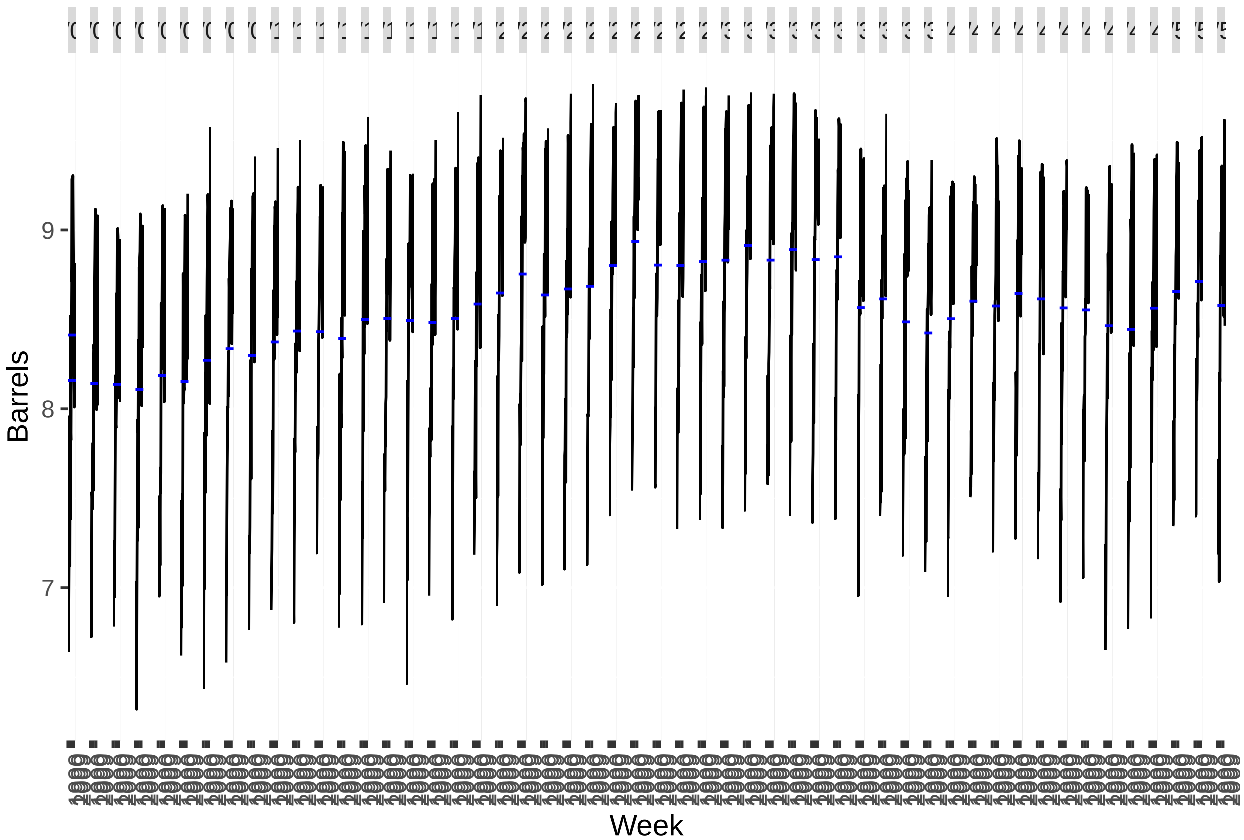
```
ts_barrels <- us_gasoline %>%
  select(Week, Barrels)

autoplot(ts_barrels, .vars = Barrels)
```



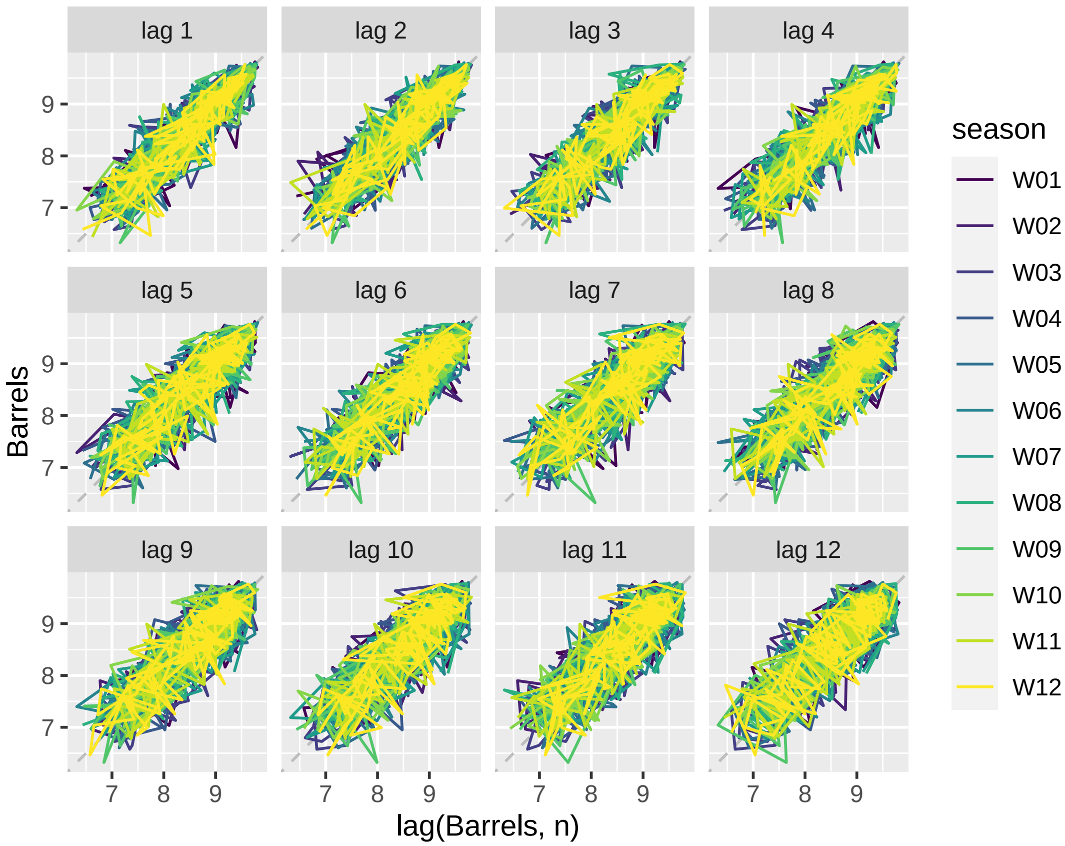

```r
gg_season(ts_barrels, y = Barrels)
```



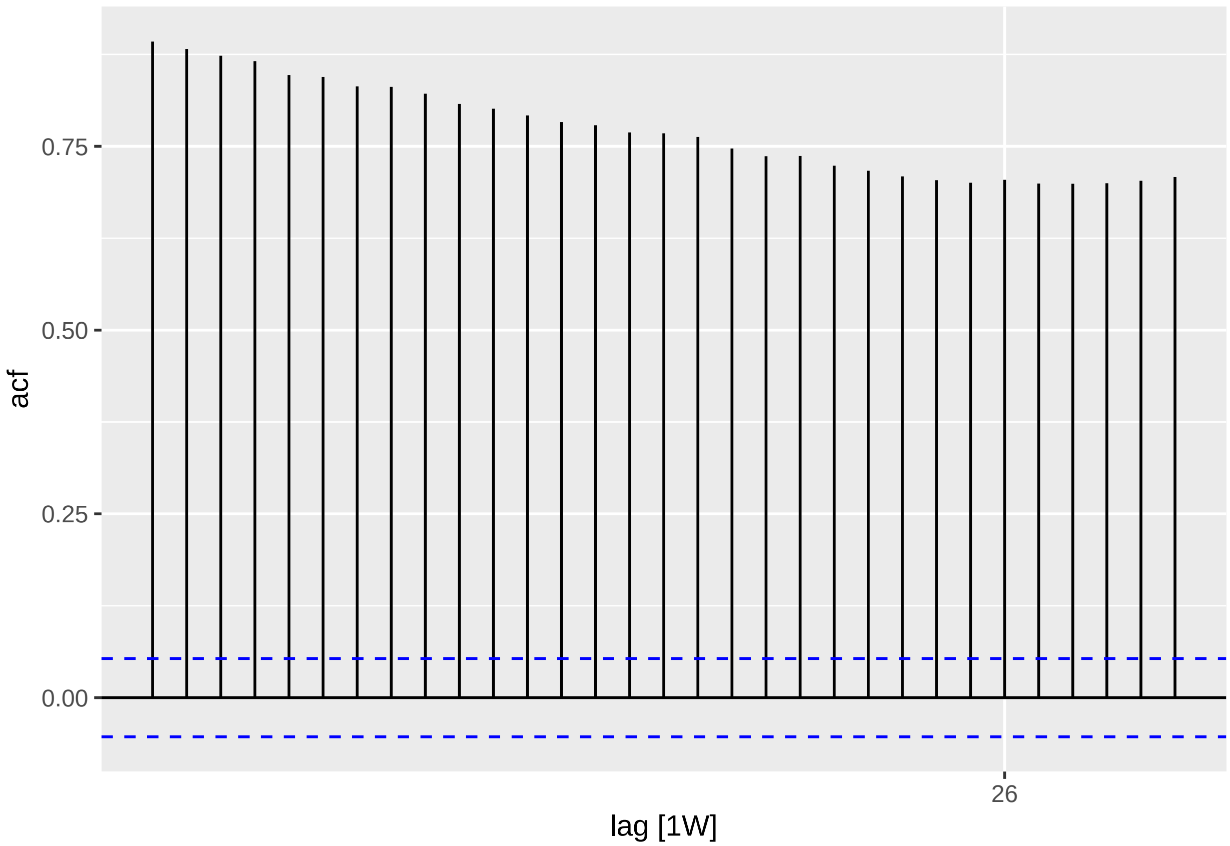

```r
gg_subseries(ts_barrels, y = Barrels)
```




```r
gg_lag(ts_barrels, y = Barrels, period = 12, lags = 1:12)
```




```
ACF(ts_barrels, y = Barrels) %>%
  autoplot()
```


0.75
0.50
0.25
0.00
acf
26
lag [1W]