Data 624 2.10 Exercises_HW1

Jiaxin Zheng

2025-02-09

2.10 Exercises:

```
library(fpp3)
```

```
## Warning: package 'fpp3' was built under R version 4.4.2
```

```
## Registered S3 method overwritten by 'tsibble':
##   method               from
##   as_tibble.grouped_df dplyr
```

```
## -- Attaching packages --------------------------------------- fpp3 1.0.1 --
```

```
## v tibble      3.2.1     v tsibble     1.1.6
## v dplyr       1.1.4     v tsibbledata 0.4.1
## v tidyr       1.3.1     v feasts      0.4.1
## v lubridate   1.9.4     v fable       0.4.1
## v ggplot2     3.5.1
```

```
## Warning: package 'dplyr' was built under R version 4.4.2
```

```
## Warning: package 'ggplot2' was built under R version 4.4.2
```

```
## Warning: package 'tsibbledata' was built under R version 4.4.2
```

```
## Warning: package 'feasts' was built under R version 4.4.2
```

```
## Warning: package 'fabletools' was built under R version 4.4.2
```

```
## Warning: package 'fable' was built under R version 4.4.2
```

```
## -- Conflicts ------------------------------------------ fpp3_conflicts --
## x lubridate::date()    masks base::date()
## x dplyr::filter()      masks stats::filter()
## x tsibble::intersect() masks base::intersect()
## x tsibble::interval()  masks lubridate::interval()
## x dplyr::lag()         masks stats::lag()
## x tsibble::setdiff()   masks base::setdiff()
## x tsibble::union()     masks base::union()
```

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

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

1.A & B:

```
help('aus_production')
```

```
## starting httpd help server ... done
```

```
data('aus_production')
# Bricks from aus_production: Clay brick production in millions of bricks. The time series is from 1956
```

```
help('pelt')
# Lynx from pelt: The number of Canadian Lynx pelts traded. The time series is from 1845 to 1935.
```

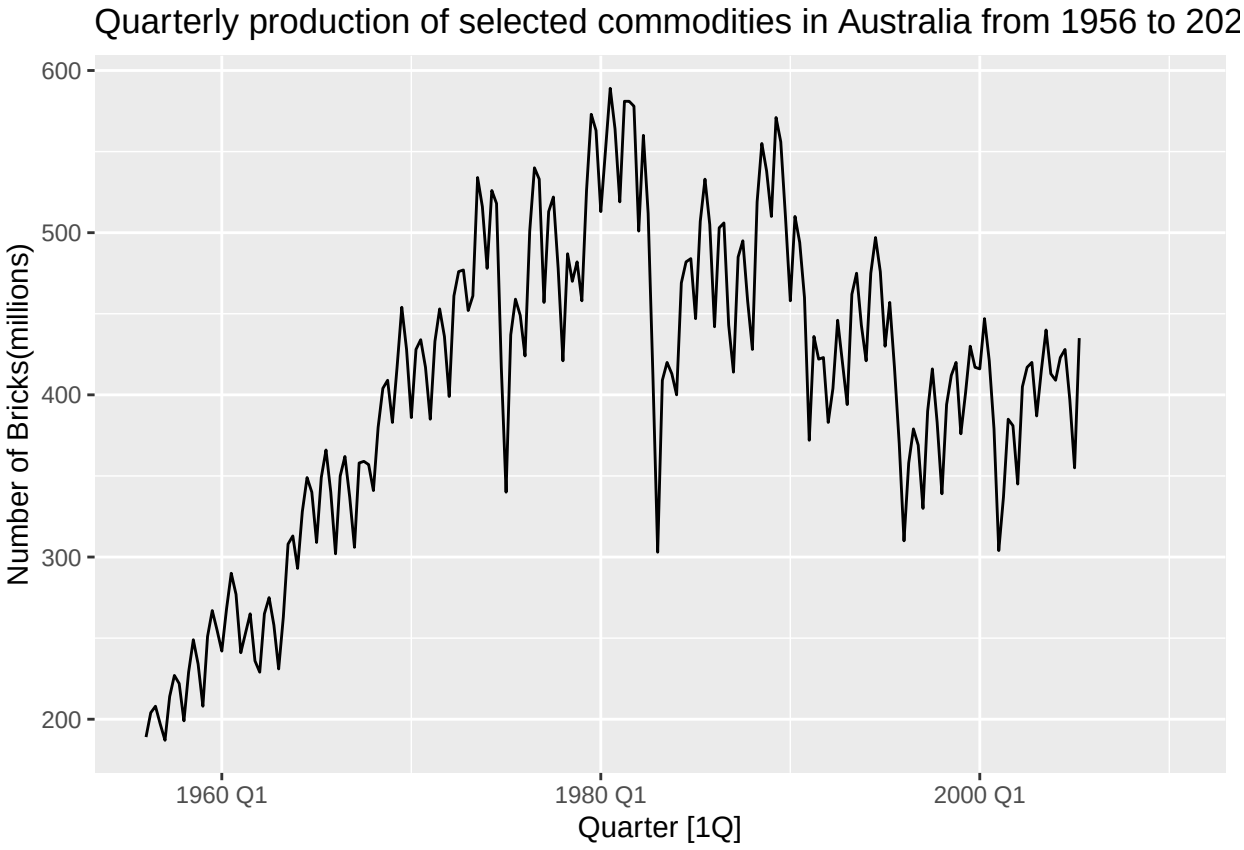
```
help('gafa_stock')
# Close from gafa_stock: The closing price for the stock. The time series is from 2014 to 2018.
```

```
help('vic_elec')
data('vic_elec')
# Demand from vic_elec: vic_elec is a half-hourly tsibble with three values: 1. Demand, 2. Temperature,
```

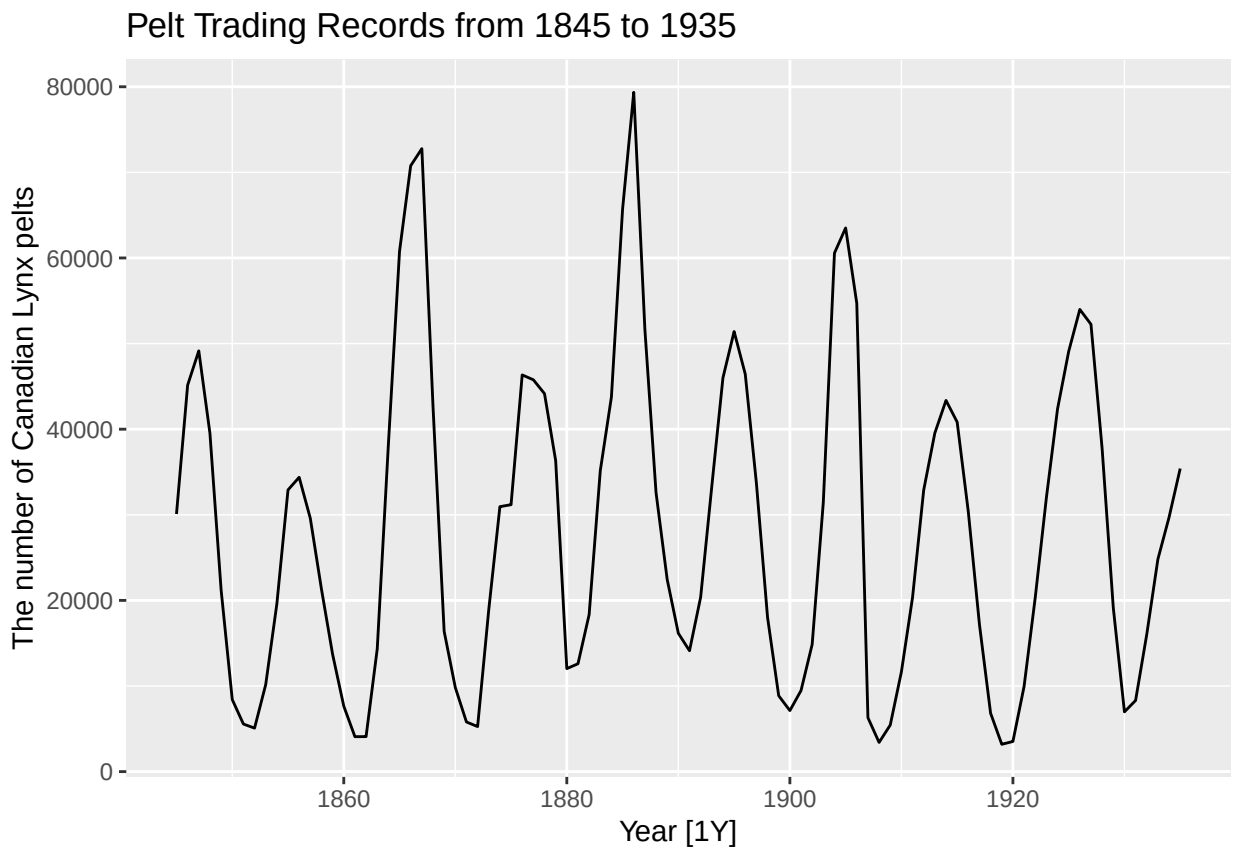
1.C & D:

```
autoplot(aus_production, Bricks) +
  labs (y = "Number of Bricks(millions)", title = "Quarterly production of selected commodities in Aust
```

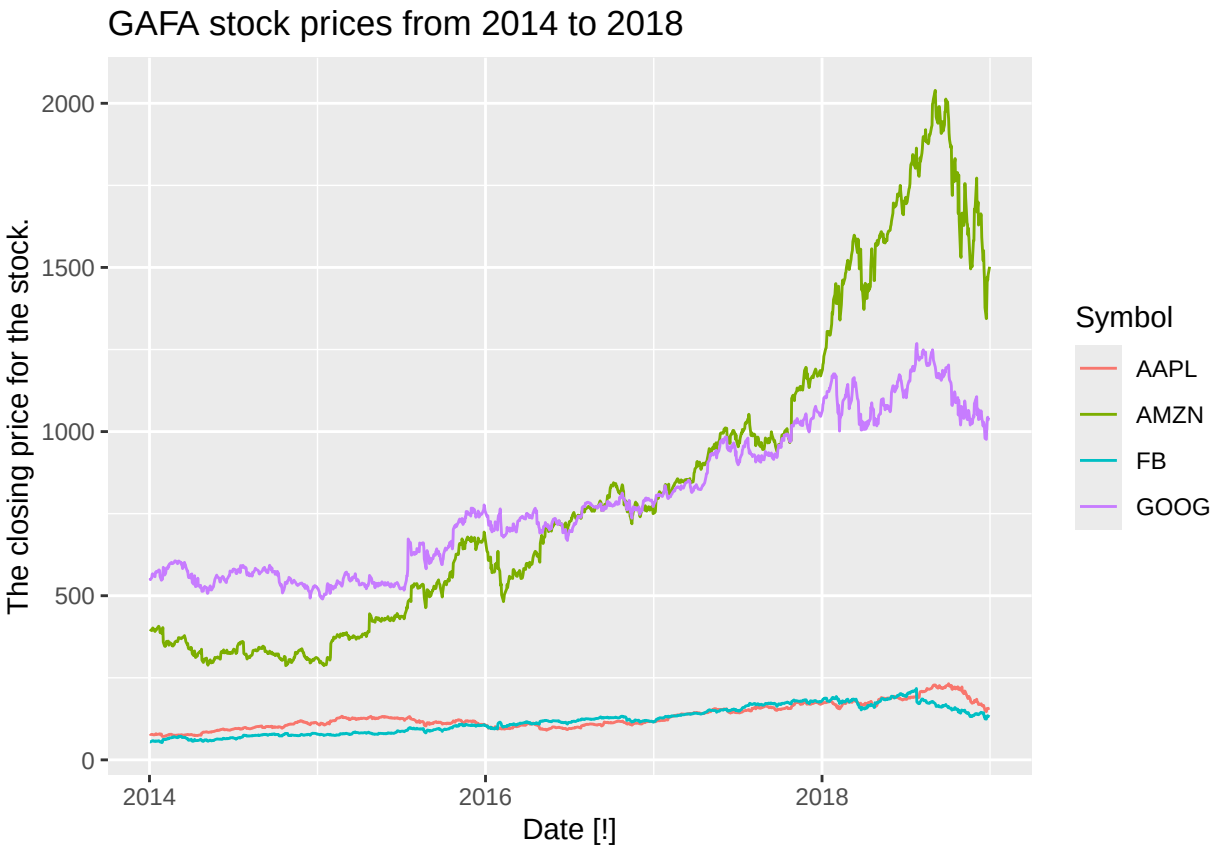
```
## Warning: Removed 20 rows containing missing values or values outside the scale range
## (`geom_line()`).
```



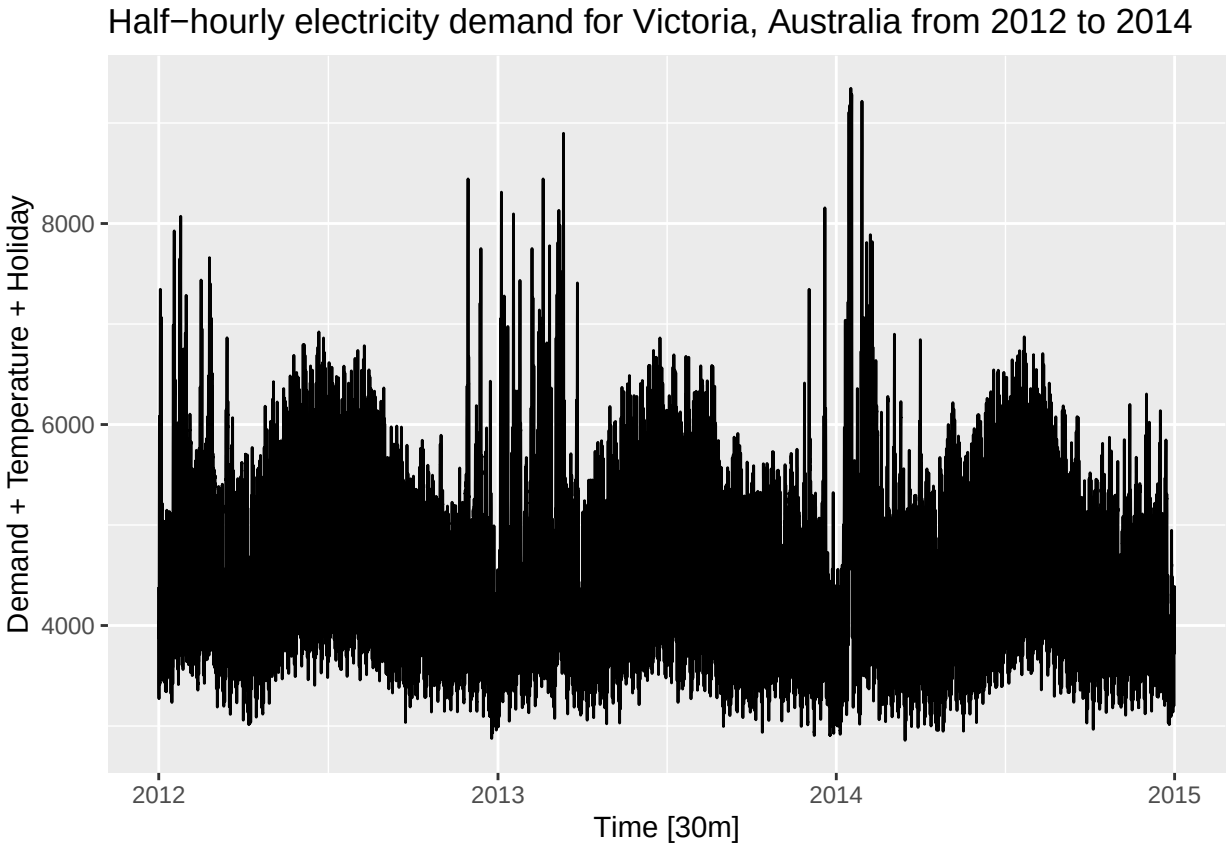

```
autoplot(pelt, Lynx) +
  labs (y = "The number of Canadian Lynx pelts", title = "Pelt Trading Records from 1845 to 1935")
```




```
autoplot(gafa_stock, Close) +
  labs (y = "The closing price for the stock.", title = "GAFA stock prices from 2014 to 2018")
```




```
autoplot(vic_elec, Demand) +
  labs (y = "Demand + Temperature + Holiday", title = "Half-hourly electricity demand for Victoria, Aus
```



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

```
head(gafa_stock)
```

```
## # A tsibble: 6 x 8 [!]
## # Key:       Symbol [1]
##   Symbol Date        Open  High   Low Close Adj_Close    Volume
##   <chr>  <date>     <dbl> <dbl> <dbl> <dbl>     <dbl>     <dbl>
## 1 AAPL   2014-01-02  79.4  79.6  78.9  79.0      67.0  58671200
## 2 AAPL   2014-01-03  79.0  79.1  77.2  77.3      65.5  98116900
## 3 AAPL   2014-01-06  76.8  78.1  76.2  77.7      65.9 103152700
## 4 AAPL   2014-01-07  77.8  78.0  76.8  77.1      65.4  79302300
## 5 AAPL   2014-01-08  77.0  77.9  77.0  77.6      65.8  64632400
## 6 AAPL   2014-01-09  78.1  78.1  76.5  76.6      65.0  69787200
```

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

```
## # A tsibble: 4 x 8 [!]
## # Key:       Symbol [4]
## # Groups:    Symbol [4]
##   Symbol Date        Open  High   Low Close Adj_Close   Volume
##   <chr>  <date>     <dbl> <dbl> <dbl> <dbl>     <dbl>    <dbl>
## 1 AAPL   2018-10-03  230.  233.  230.  232.      230. 28654800
```

```
## 2 AMZN   2018-09-04 2026. 2050. 2013  2040.     2040.  5721100
## 3 FB     2018-07-25  216.  219.  214.  218.      218. 58954200
## 4 GOOG   2018-07-26 1251  1270. 1249. 1268.     1268.  2405600
```

3. 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.

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

```
tute1 <- read.csv("https://raw.githubusercontent.com/Jennyjjxxzz/Data-624_HW1/refs/heads/main/tute1.csv
head(tute1)
```

```
##      Quarter  Sales AdBudget   GDP
## 1 1981-03-01 1020.2    659.2 251.8
## 2 1981-06-01  889.2    589.0 290.9
## 3 1981-09-01  795.0    512.5 290.8
## 4 1981-12-01 1003.9    614.1 292.4
## 5 1982-03-01 1057.7    647.2 279.1
## 6 1982-06-01  944.4    602.0 254.0
```

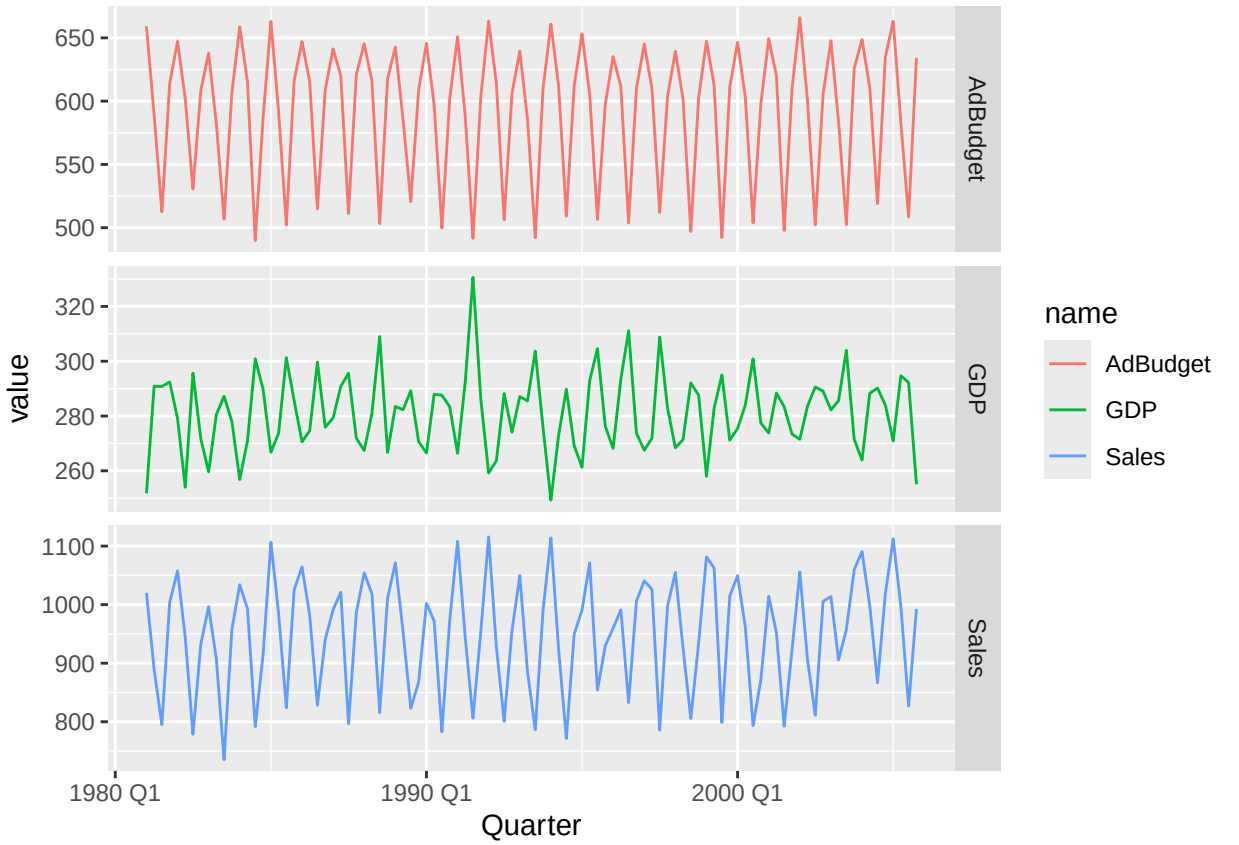
3.b. Convert the data to time series

```
mytimeseries <- tute1 |>
  mutate(Quarter = yearquarter(Quarter)) |>
  as_tsibble(index = Quarter)
```

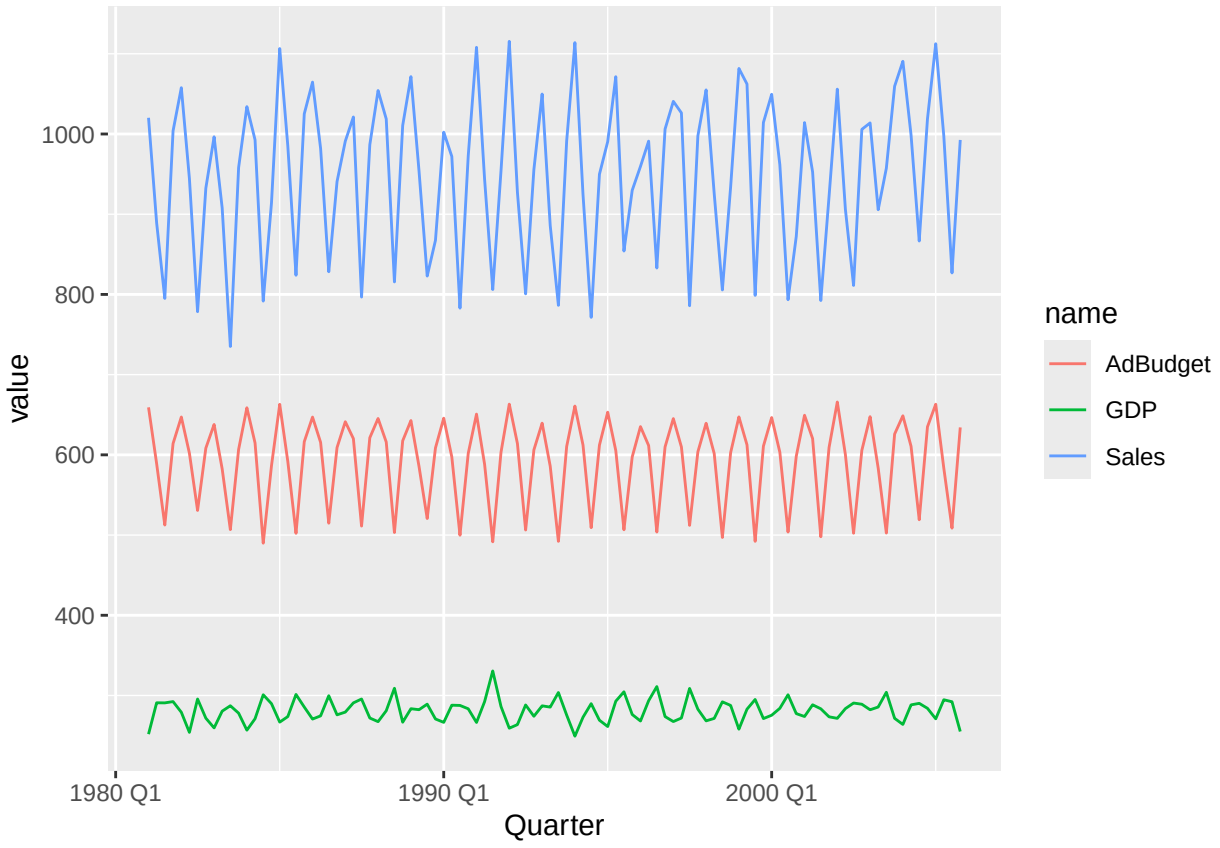
3.c. Construct time series plots of each of the three series

- If we don't include facet_grid(), the name will mix together.

```
mytimeseries |>
  pivot_longer(-Quarter) |>
  ggplot(aes(x = Quarter, y = value, colour = name)) +
  geom_line() +
  facet_grid(name ~ ., scales = "free_y")
```




```
mytimeseries |>
  pivot_longer(-Quarter) |>
  ggplot(aes(x = Quarter, y = value, colour = name)) +
  geom_line()
```



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).

4.a.

```
library(USgas)
```

```
## Warning: package 'USgas' was built under R version 4.4.2
```

4.b.

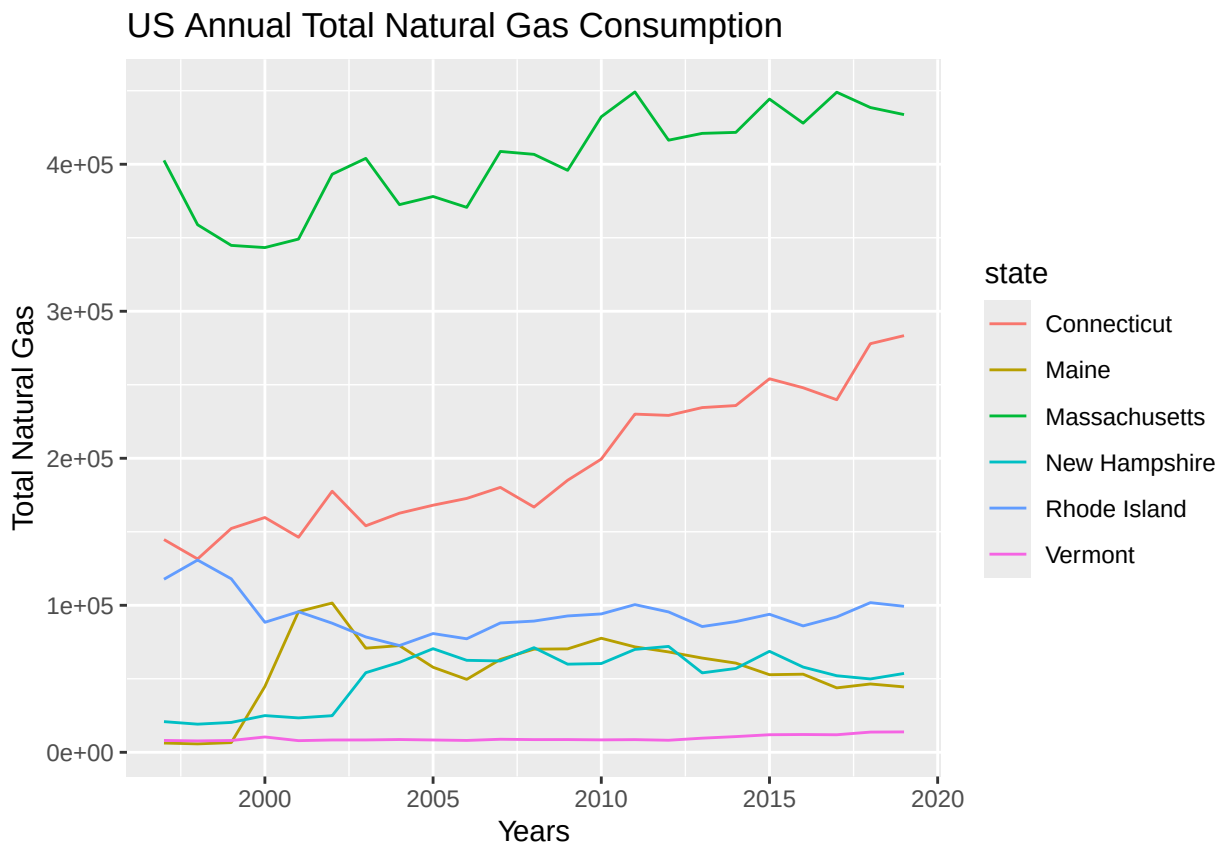
```
us_total <- us_total %>%
  tsibble(index = year,key = state)

us_total
```

```
## # A tsibble: 1,266 x 3 [1Y]
## # Key:       state [53]
##     year state        y
##    <int> <chr>    <int>
##  1  1997 Alabama 324158
##  2  1998 Alabama 329134
##  3  1999 Alabama 337270
##  4  2000 Alabama 353614
##  5  2001 Alabama 332693
##  6  2002 Alabama 379343
##  7  2003 Alabama 350345
##  8  2004 Alabama 382367
##  9  2005 Alabama 353156
## 10  2006 Alabama 391093
## # i 1,256 more rows
```

4.c.

```
us_total %>%
  filter(state %in% c('Maine', 'Vermont', 'New Hampshire', 'Massachusetts', 'Connecticut', 'Rhode Islan
  ggplot(aes(x = year, y = y, colour = state)) +
  geom_line()+
  labs (y = "Total Natural Gas", x = "Years", title = "US Annual Total Natural Gas Consumption")
```



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.
- c.Find what combination of Region and Purpose had the maximum number of overnight trips on average.
- d.Create a new tsibble which combines the Purposes and Regions, and just has total trips by State.

5.a.

```
tourism <- readxl::read_excel("tourism.xlsx")
head(tourism)
```

```
## # A tibble: 6 x 5
##   Quarter    Region   State           Purpose  Trips
##   <chr>      <chr>    <chr>           <chr>    <dbl>
## 1 1998-01-01 Adelaide South Australia Business  135.
## 2 1998-04-01 Adelaide South Australia Business  110.
## 3 1998-07-01 Adelaide South Australia Business  166.
## 4 1998-10-01 Adelaide South Australia Business  127.
## 5 1999-01-01 Adelaide South Australia Business  137.
## 6 1999-04-01 Adelaide South Australia Business  200.
```

5.b.

```
tibble_tourism <- tourism %>%
  mutate(Quarter = yearquarter(Quarter)) %>%
  as_tsibble(key = c("Region", "State", "Purpose", "Trips"),
             index = Quarter)

tibble_tourism
```

```
## # A tsibble: 24,320 x 5 [1Q]
## # Key:       Region, State, Purpose, Trips [22,871]
##    Quarter Region   State           Purpose  Trips
##      <qtr> <chr>    <chr>           <chr>    <dbl>
##  1 2010 Q1 Adelaide South Australia Business  68.7
##  2 2005 Q2 Adelaide South Australia Business  73.3
##  3 2013 Q2 Adelaide South Australia Business 101.
##  4 2001 Q4 Adelaide South Australia Business 101.
##  5 2013 Q1 Adelaide South Australia Business 102.
##  6 2006 Q4 Adelaide South Australia Business 107.
##  7 2011 Q1 Adelaide South Australia Business 110.
##  8 1998 Q2 Adelaide South Australia Business 110.
##  9 2009 Q1 Adelaide South Australia Business 114.
## 10 2010 Q3 Adelaide South Australia Business 121.
## # i 24,310 more rows
```

5.c.

```
tibble_tourism2 <- tibble_tourism %>%
  select(Region, Purpose, Trips) %>%
  group_by(Region, Purpose) %>%
  summarise(Avg_Trips = mean(Trips)) %>%
  filter(Avg_Trips == max(Avg_Trips))%>%
  arrange(desc(Avg_Trips))

tibble_tourism2
```

```
## # A tsibble: 76 x 4 [1Q]
## # Key:       Region, Purpose [76]
## # Groups:    Region [76]
##    Region                 Purpose  Quarter Avg_Trips
##    <chr>                  <chr>      <qtr>     <dbl>
##  1 Melbourne              Visiting 2017 Q4      985.
##  2 Sydney                 Business 2001 Q4      948.
##  3 South Coast            Holiday  1998 Q1      915.
##  4 North Coast NSW        Holiday  2016 Q1      906.
##  5 Brisbane               Visiting 2016 Q4      796.
##  6 Gold Coast             Holiday  2002 Q1      711.
##  7 Sunshine Coast         Holiday  2005 Q1      617.
##  8 Australia's South West Holiday  2016 Q1      612.
##  9 Great Ocean Road       Holiday  1998 Q1      548.
## 10 Experience Perth       Visiting 2016 Q1      538.
## # i 66 more rows
```

5.d.

```
tibble_tourism3 <- tourism %>%
  group_by(State) %>%
  summarise(Total = sum(Trips))
tibble_tourism3
```

```
## # A tibble: 8 x 2
##   State                Total
##   <chr>                <dbl>
## 1 ACT                 41007.
## 2 New South Wales    557367.
## 3 Northern Territory  28614.
## 4 Queensland         386643.
## 5 South Australia    118151.
## 6 Tasmania            54137.
## 7 Victoria           390463.
## 8 Western Australia  147820.
```

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.

us_employment Data:

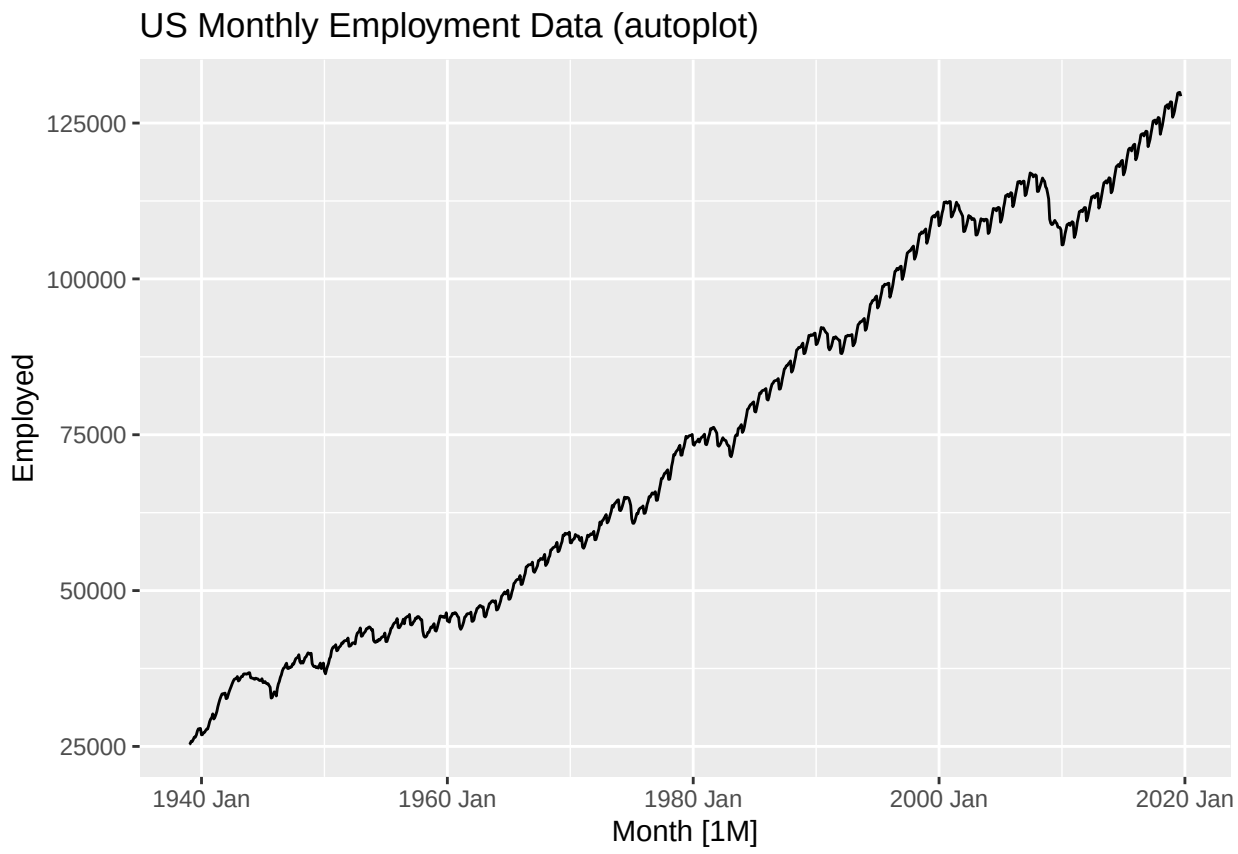
- a.Can you spot any seasonality, cyclicity and trend?
- answer a:The trend of the US employment from 1939 to 2019 are increase, the trend is going upward.
- b.What do you learn about the series?
- answer b:The series shows the increase in employment throughout different years and months. The growth has been consistent.
- c.What can you say about the seasonal patterns?
- answer c:The seasonal patterns does not show particular season with big affect.
- d.Can you identify any unusual years?
- answer d:Around year of 2021, there is a small dip, may be the economic recession.

```
?us_employment
```

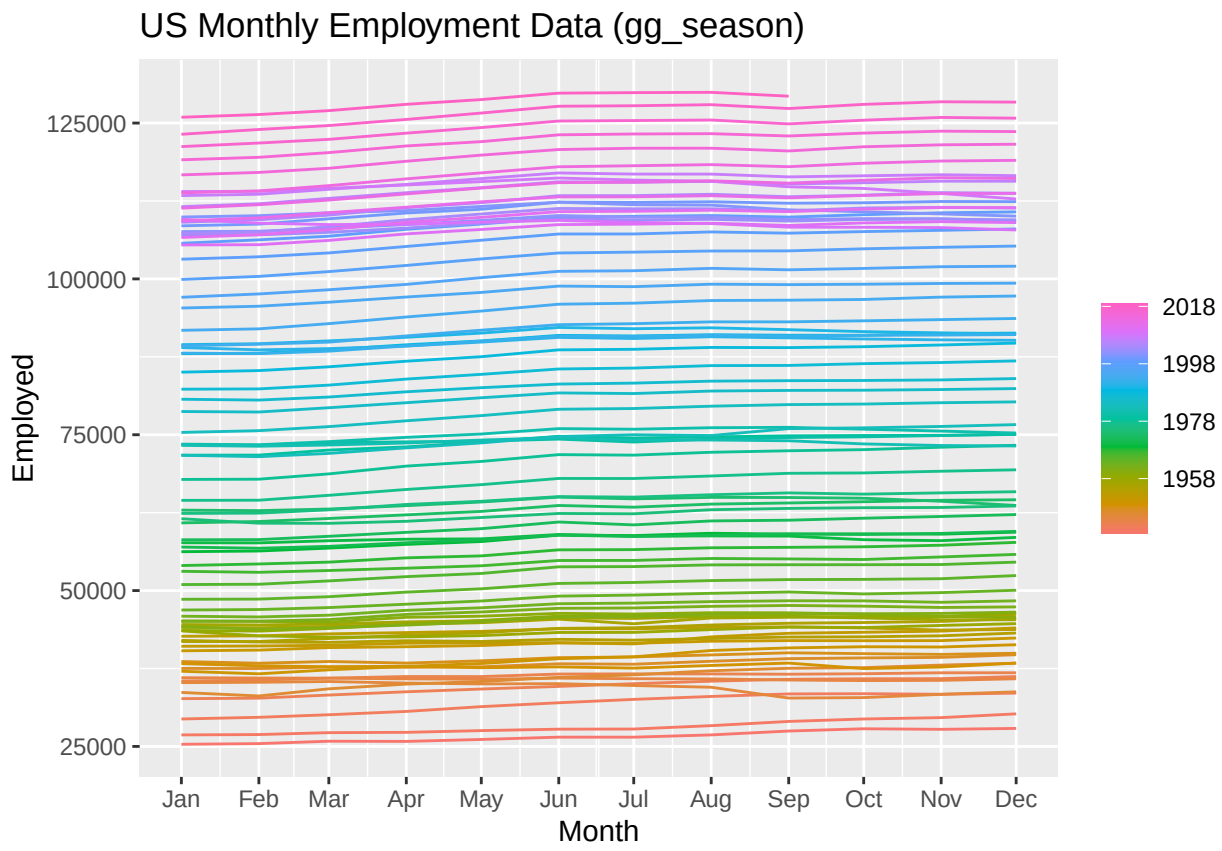
```
us_employment
```

```
## # A tsibble: 143,412 x 4 [1M]
## # Key:       Series_ID [148]
##       Month Series_ID     Title          Employed
##       <mth> <chr>         <chr>             <dbl>
##  1 1939 Jan CEU0500000001 Total Private    25338
##  2 1939 Feb CEU0500000001 Total Private    25447
##  3 1939 Mar CEU0500000001 Total Private    25833
##  4 1939 Apr CEU0500000001 Total Private    25801
##  5 1939 May CEU0500000001 Total Private    26113
##  6 1939 Jun CEU0500000001 Total Private    26485
##  7 1939 Jul CEU0500000001 Total Private    26481
##  8 1939 Aug CEU0500000001 Total Private    26848
##  9 1939 Sep CEU0500000001 Total Private    27468
## 10 1939 Oct CEU0500000001 Total Private    27830
## # i 143,402 more rows
```

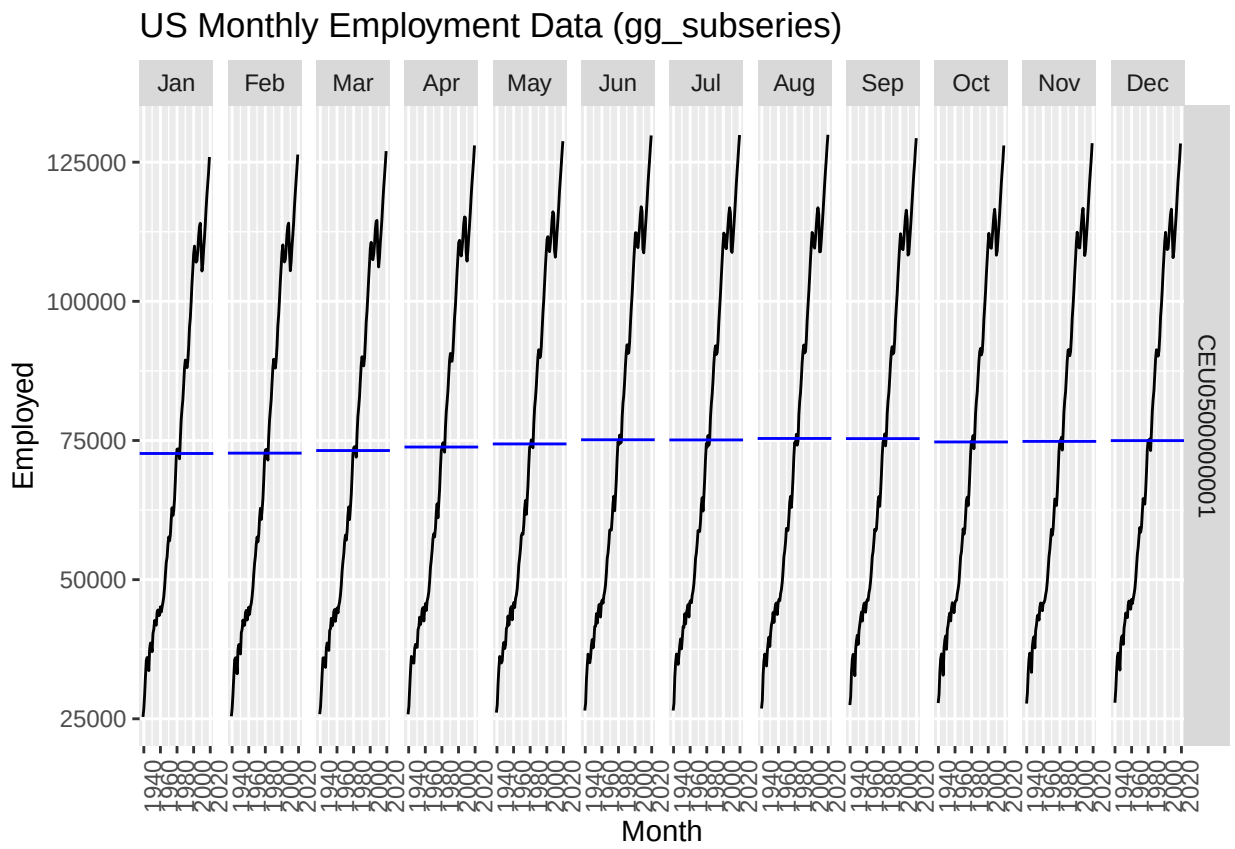
```
us_employment %>%
  filter(Title == "Total Private") %>%
autoplot(Employed)+
  labs(title = "US Monthly Employment Data (autoplot)")
```




```
us_employment %>%
  filter(Title == "Total Private") %>%
gg_season(Employed)+
  labs(title = "US Monthly Employment Data (gg_season)")
```

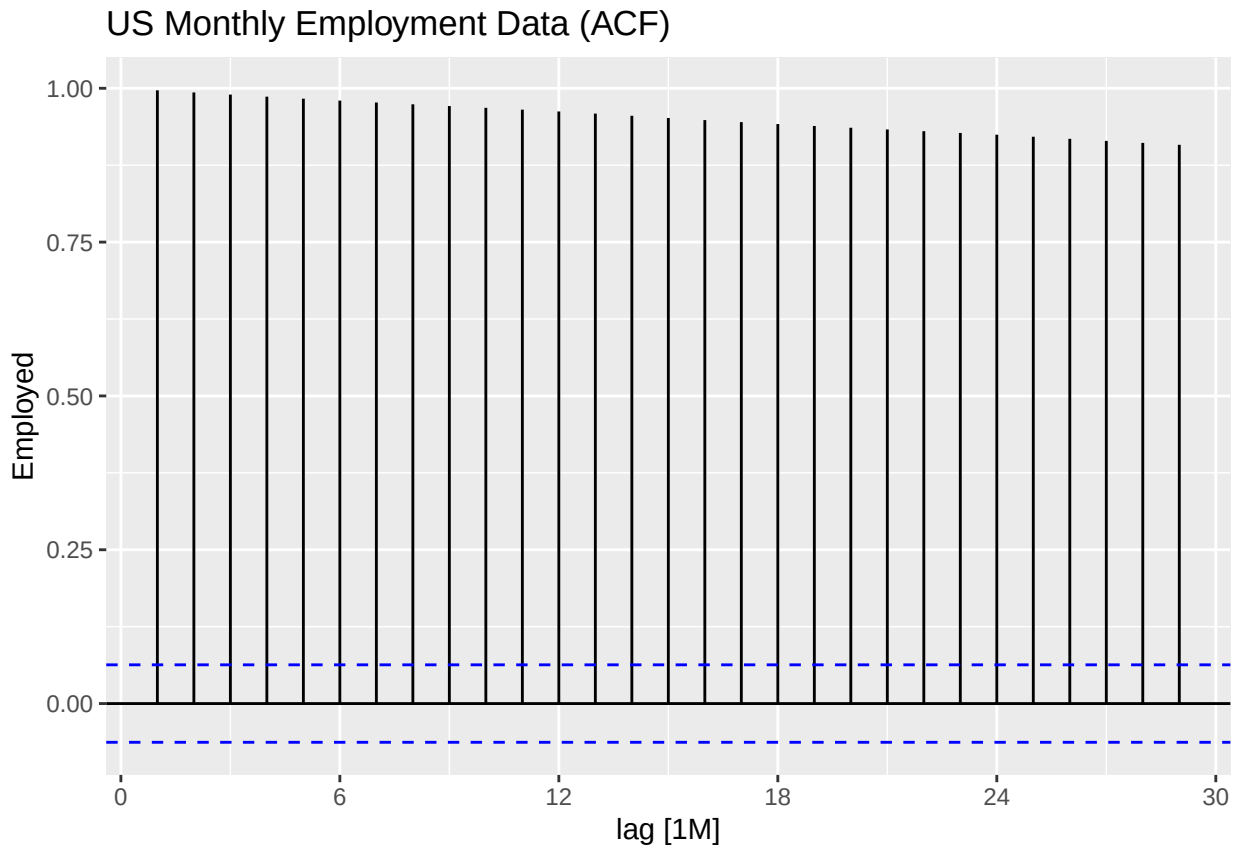



```
us_employment %>%
  filter(Title == "Total Private") %>%
gg_subseries(Employed)+
  labs(title = "US Monthly Employment Data (gg_subseries)")
```




```
us_employment %>%
  filter(Title == "Total Private") %>%
gg_lag(Employed, geom = "point")+
  labs(title = "US Monthly Employment Data (gg_lag)")
```


```
us_employment %>%
  filter(Title == "Total Private") %>%
  ACF(Employed) %>%
autoplot() +
  labs(y = "Employed", title="US Monthly Employment Data (ACF)")
```



aus_production Data:

- a.Can you spot any seasonality, cyclicity and trend?
- answer a:Even there is a lot of spikes and dips, but there is a positive upward trend when you see in long term.
- b.What do you learn about the series?
- answer b:gg_subseries shows in Q3 is the peak season of the Bricks production.
- c.What can you say about the seasonal patterns?
- answer c:The Q1 is usually the lower season than others. The Q3 is the peak season of the Bricks production.
- d.Can you identify any unusual years?
- answer d:Around the 1980s, maybe the Q2 season, the Bricks production had a significant dip.

```
?aus_production
```

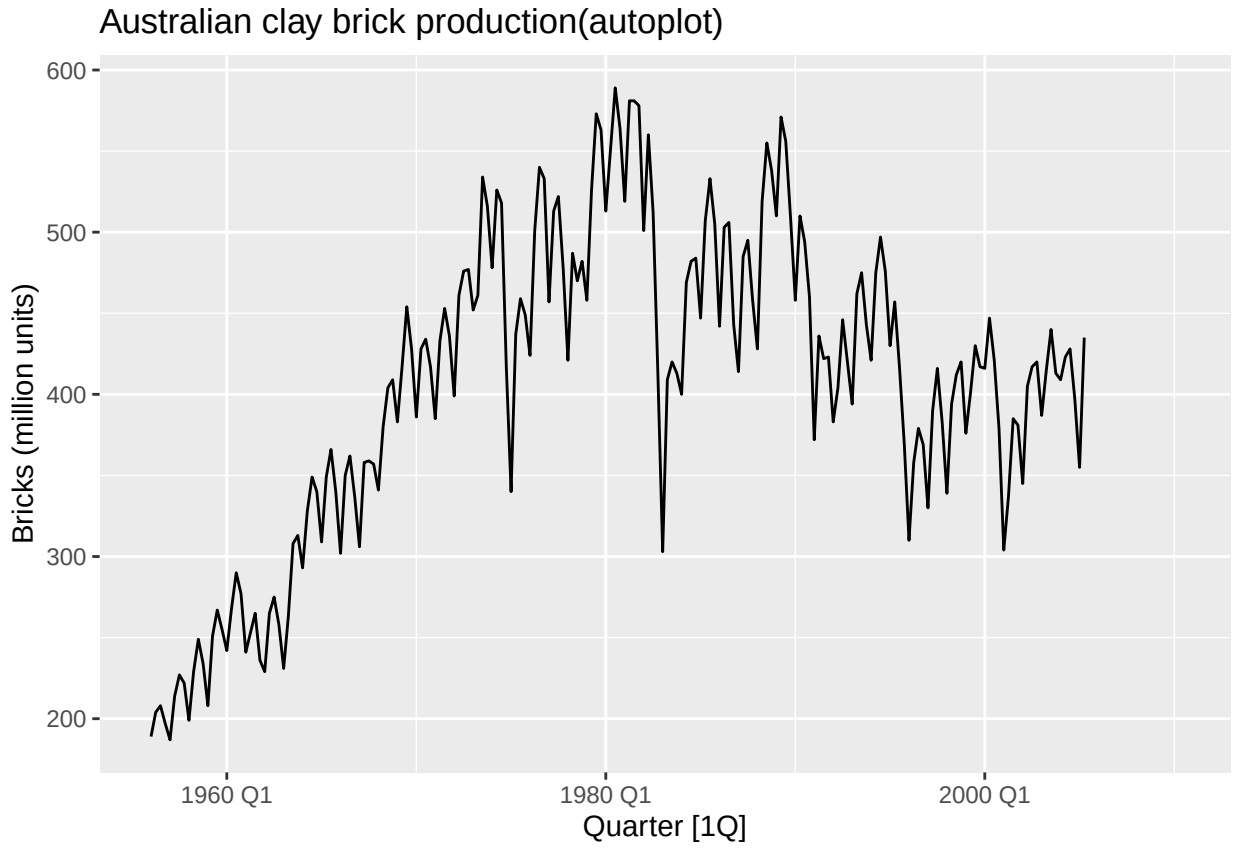
```
aus_production
```

```
## # A tsibble: 218 x 7 [1Q]
##    Quarter  Beer Tobacco Bricks Cement Electricity   Gas
##      <qtr> <dbl>   <dbl>  <dbl>  <dbl>       <dbl> <dbl>
##  1 1956 Q1   284    5225    189    465        3923     5
##  2 1956 Q2   213    5178    204    532        4436     6
##  3 1956 Q3   227    5297    208    561        4806     7
##  4 1956 Q4   308    5681    197    570        4418     6
```

```
##  5 1957 Q1   262    5577    187    529        4339     5
##  6 1957 Q2   228    5651    214    604        4811     7
##  7 1957 Q3   236    5317    227    603        5259     7
##  8 1957 Q4   320    6152    222    582        4735     6
##  9 1958 Q1   272    5758    199    554        4608     5
## 10 1958 Q2   233    5641    229    620        5196     7
## # i 208 more rows
```

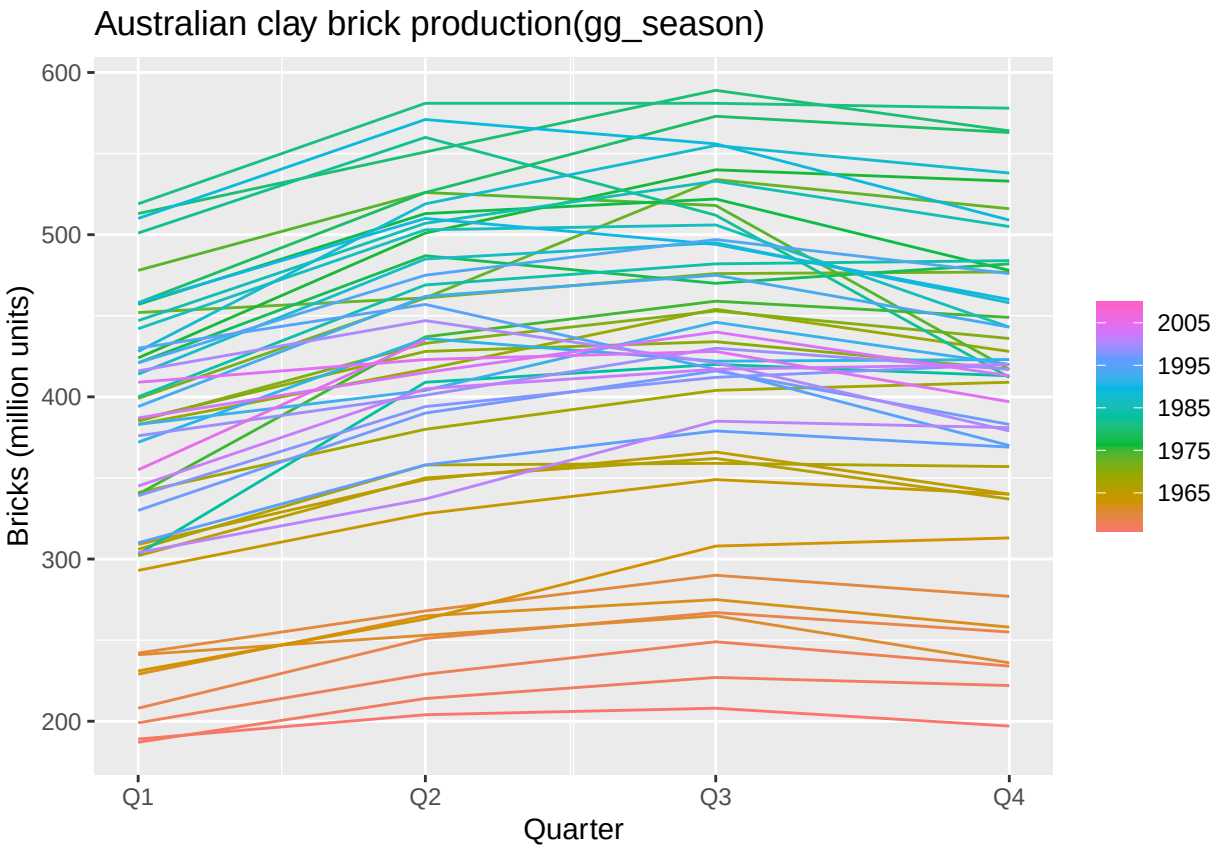
```
aus_production %>%
  autoplot(Bricks) +
  labs(y = "Bricks (million units)", title = "Australian clay brick production(autoplot)")
```

```
## Warning: Removed 20 rows containing missing values or values outside the scale range
## (`geom_line()`).
```



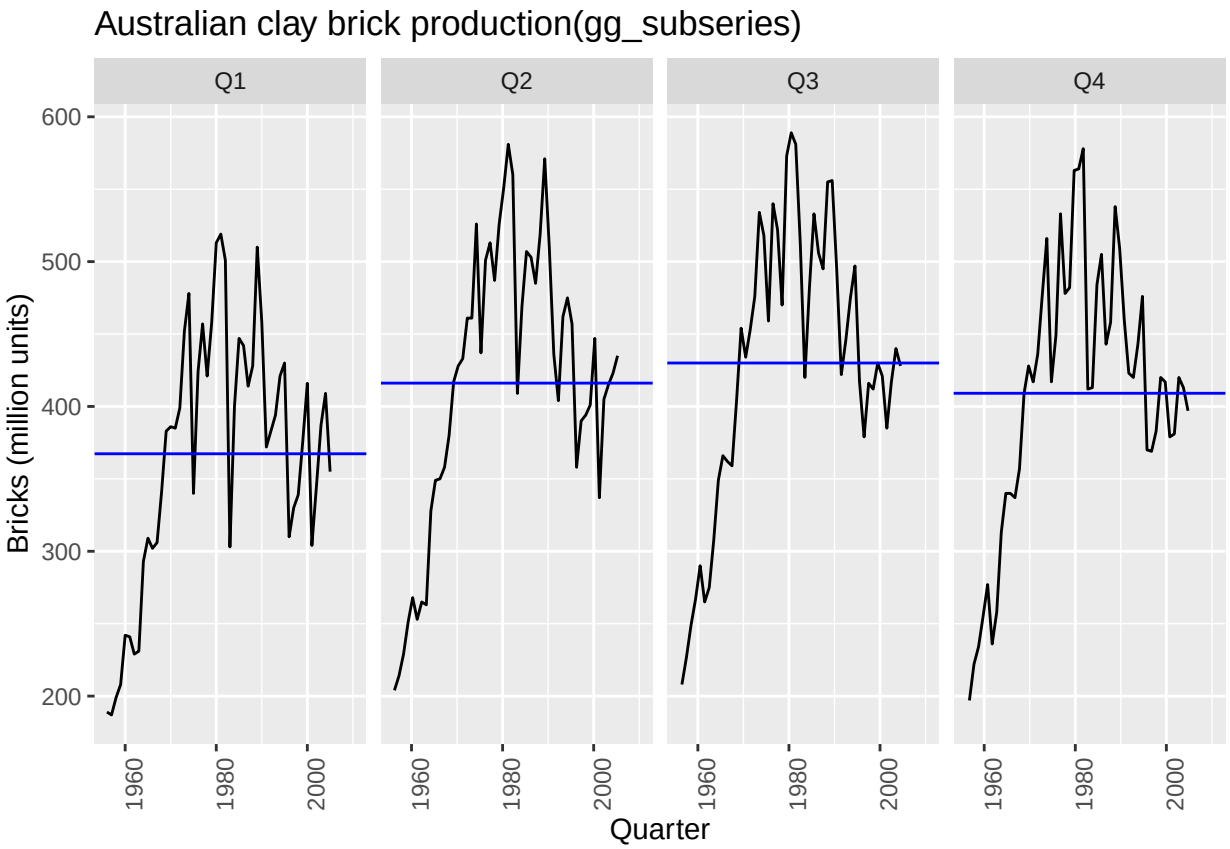

```
aus_production %>%
  gg_season(Bricks) +
  labs(y = "Bricks (million units)", title = "Australian clay brick production(gg_season)")
```

```
## Warning: Removed 20 rows containing missing values or values outside the scale range
## (`geom_line()`).
```



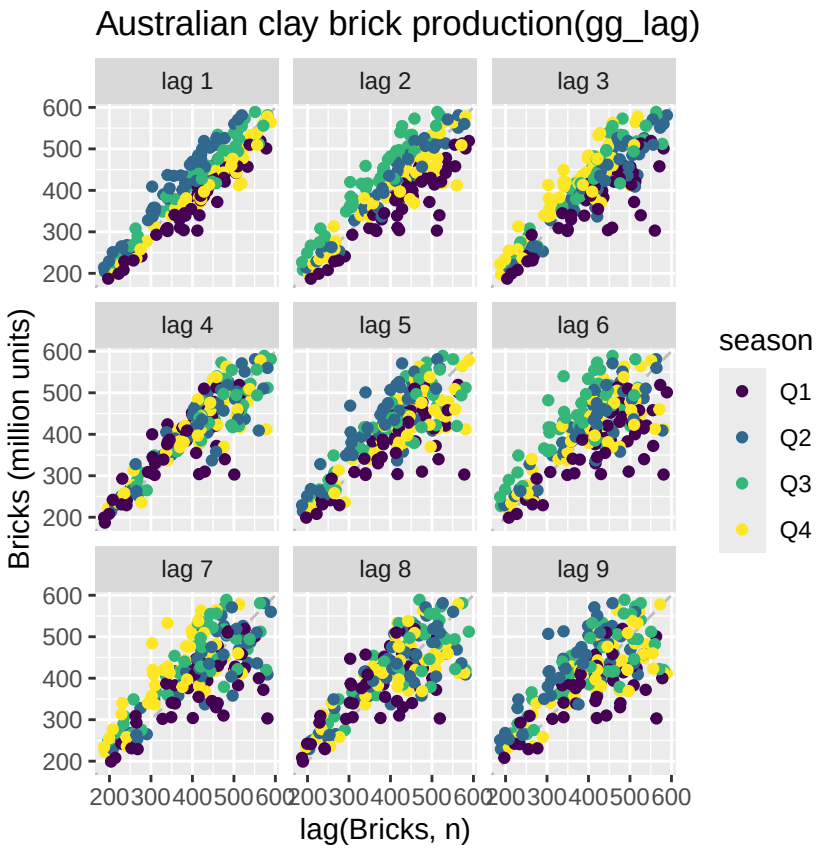

```
aus_production %>%
  gg_subseries(Bricks)+
  labs(y = "Bricks (million units)", title = "Australian clay brick production(gg_subseries)")
```

```
## Warning: Removed 5 rows containing missing values or values outside the scale range
## (`geom_line()`).
```

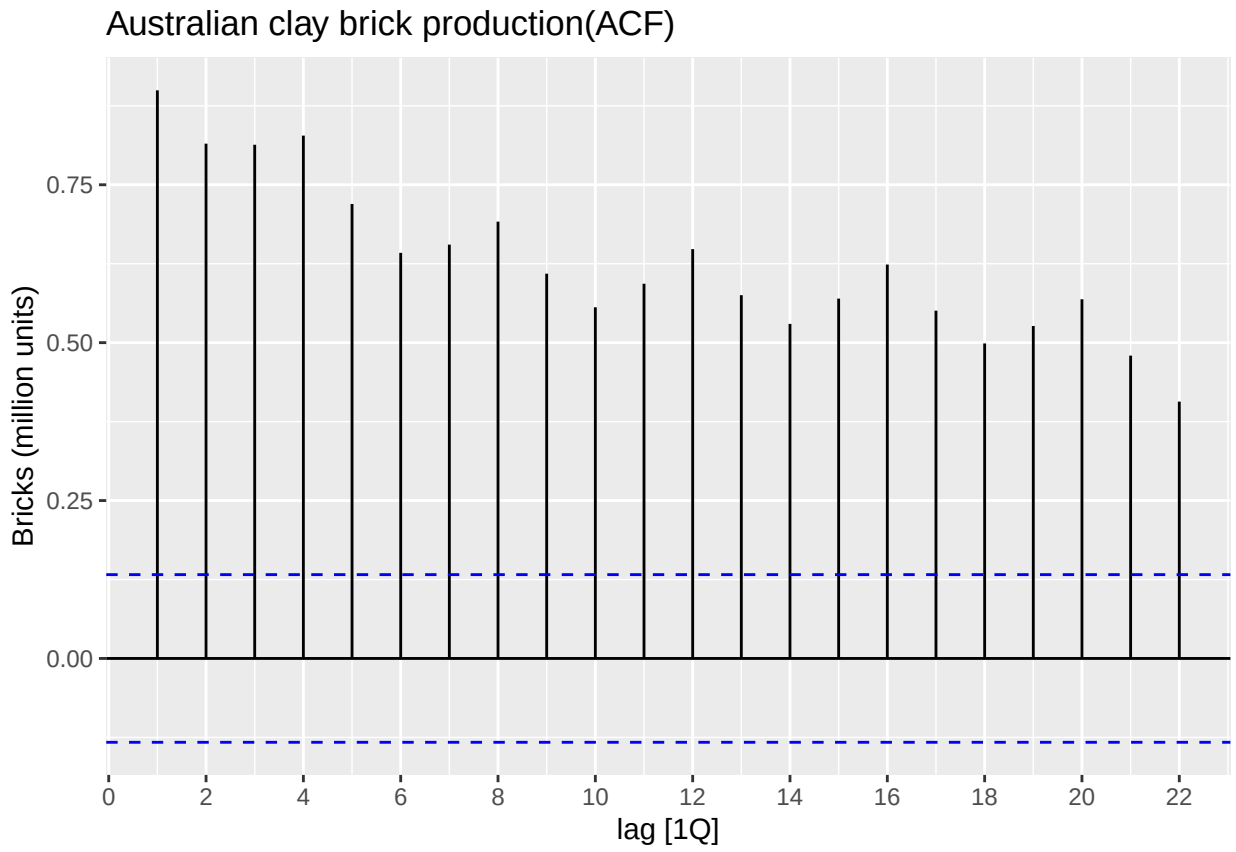



```
aus_production %>%
  gg_lag(Bricks, geom = "point")+
  labs(y = "Bricks (million units)", title = "Australian clay brick production(gg_lag)")
```

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




```
aus_production %>%
  ACF(Bricks) %>%
  autoplot()+
  labs(y = "Bricks (million units)", title = "Australian clay brick production(ACF)")
```



pelt Data:

- a.Can you spot any seasonality, cyclicity and trend?
- answer a:The data is base at annual, can't really tell the seasonality. The trend shows up and down, I think that is definitely in cyclical and seasonality for pelt trade.
- b.What do you learn about the series?
- answer b: The trend goes up and down, and varies a great deal.
- c.What can you say about the seasonal patterns?
- answer c: Can't really tell the season patterns, but seens every 5 years there was a big change.
- d.Can you identify any unusual years?
- answer d:Around 1963- 1965 the pelt trade reached to peak (maybe there was a fashion trend during that period?).

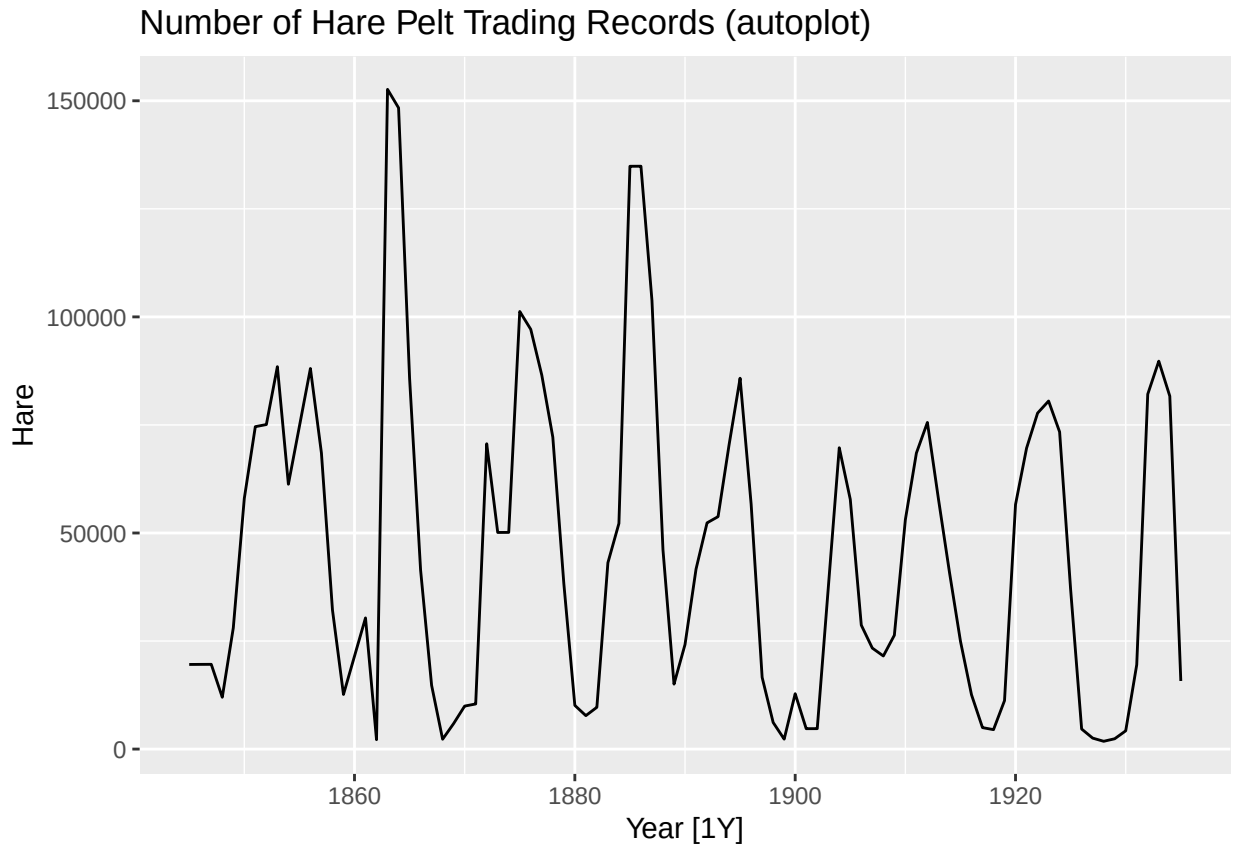
```
?pelt
```

```
pelt
```

```
## # A tsibble: 91 x 3 [1Y]
##     Year  Hare  Lynx
##    <dbl> <dbl> <dbl>
##  1  1845 19580 30090
##  2  1846 19600 45150
##  3  1847 19610 49150
##  4  1848 11990 39520
```

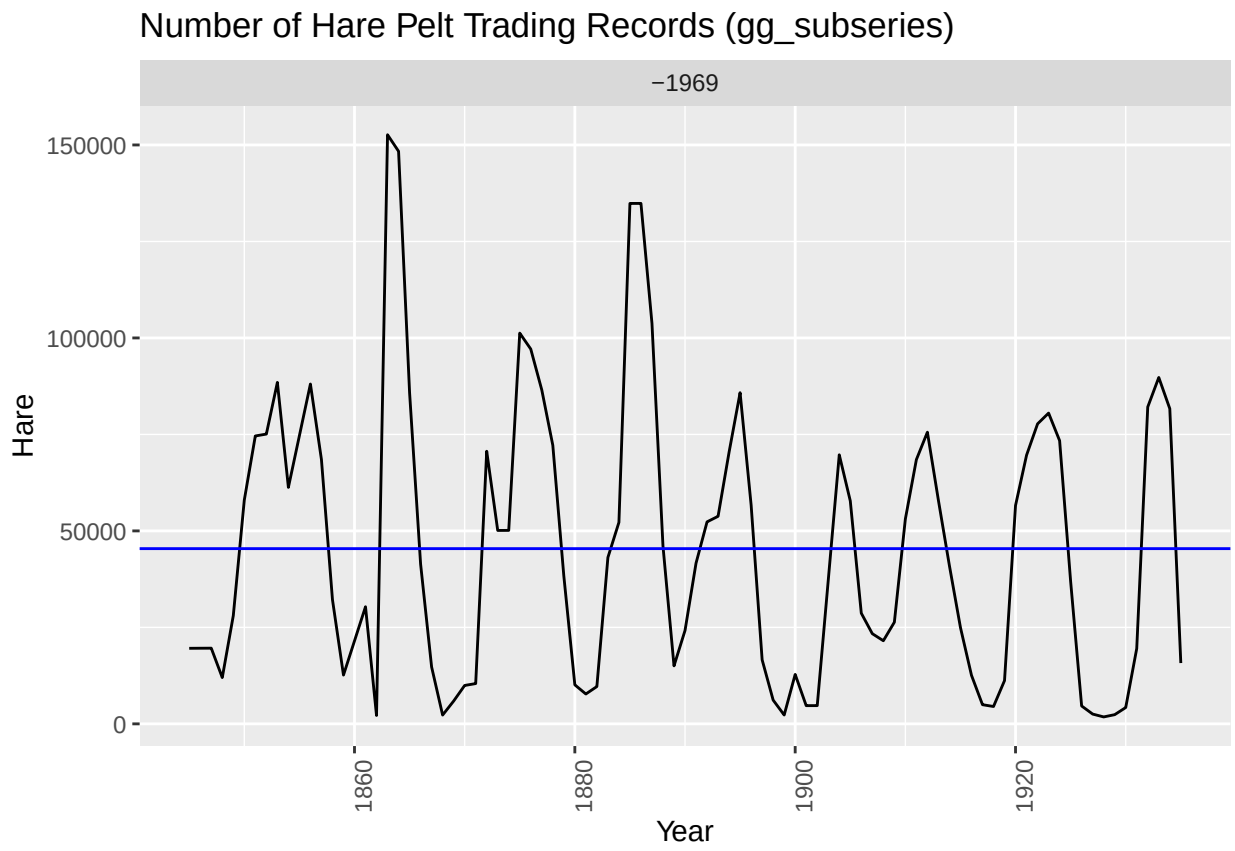
```
##  5  1849 28040 21230
##  6  1850 58000  8420
##  7  1851 74600  5560
##  8  1852 75090  5080
##  9  1853 88480 10170
## 10  1854 61280 19600
## # i 81 more rows
```

```
pelt %>%
  autoplot(Hare) +
  labs(title = "Number of Hare Pelt Trading Records (autoplot)")
```

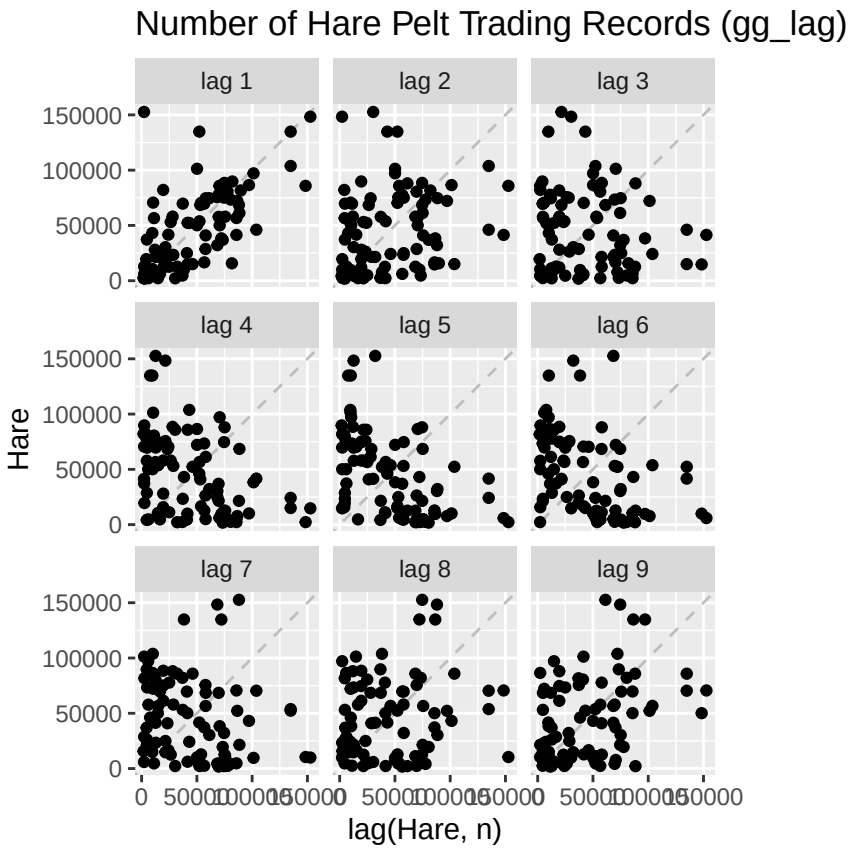



```
#Fail to plot the gg_season for Hare
#pelt %>%
  #gg_season(Hare) +
  #labs(title = "Number of Hare Pelt Trading Records (gg_season)")
```

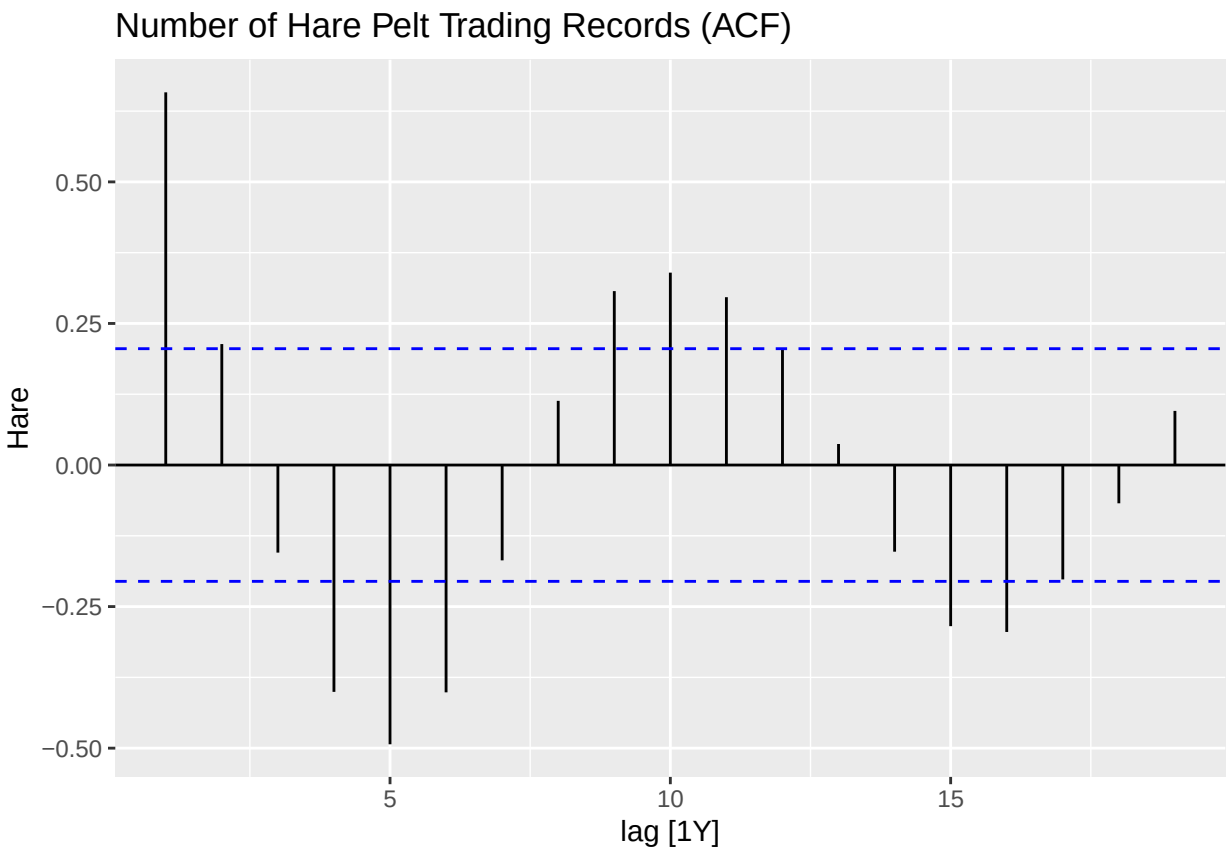
```
pelt %>%
  gg_subseries(Hare)+
  labs(title = "Number of Hare Pelt Trading Records (gg_subseries)")
```




```
pelt %>%
  gg_lag(Hare, geom = "point")+
  labs(title = "Number of Hare Pelt Trading Records (gg_lag)")
```




```
pelt %>%
  ACF(Hare) %>%
  autoplot()+
  labs(y = "Hare", title = "Number of Hare Pelt Trading Records (ACF)")
```



PBS Data:

- a.Can you spot any seasonality, cyclicity and trend?
- answer a:The trends go up and down, but it seems to be seasonality, and cyclicity.
- b.What do you learn about the series?
- answer b:The trends go up and down, but it seems to be increase.
- c.What can you say about the seasonal patterns?
- answer c:Sometimes during the spring season and the end of the year, the cost of H02 increases.
- d.Can you identify any unusual years?
- answer d:No particular year that stands out.

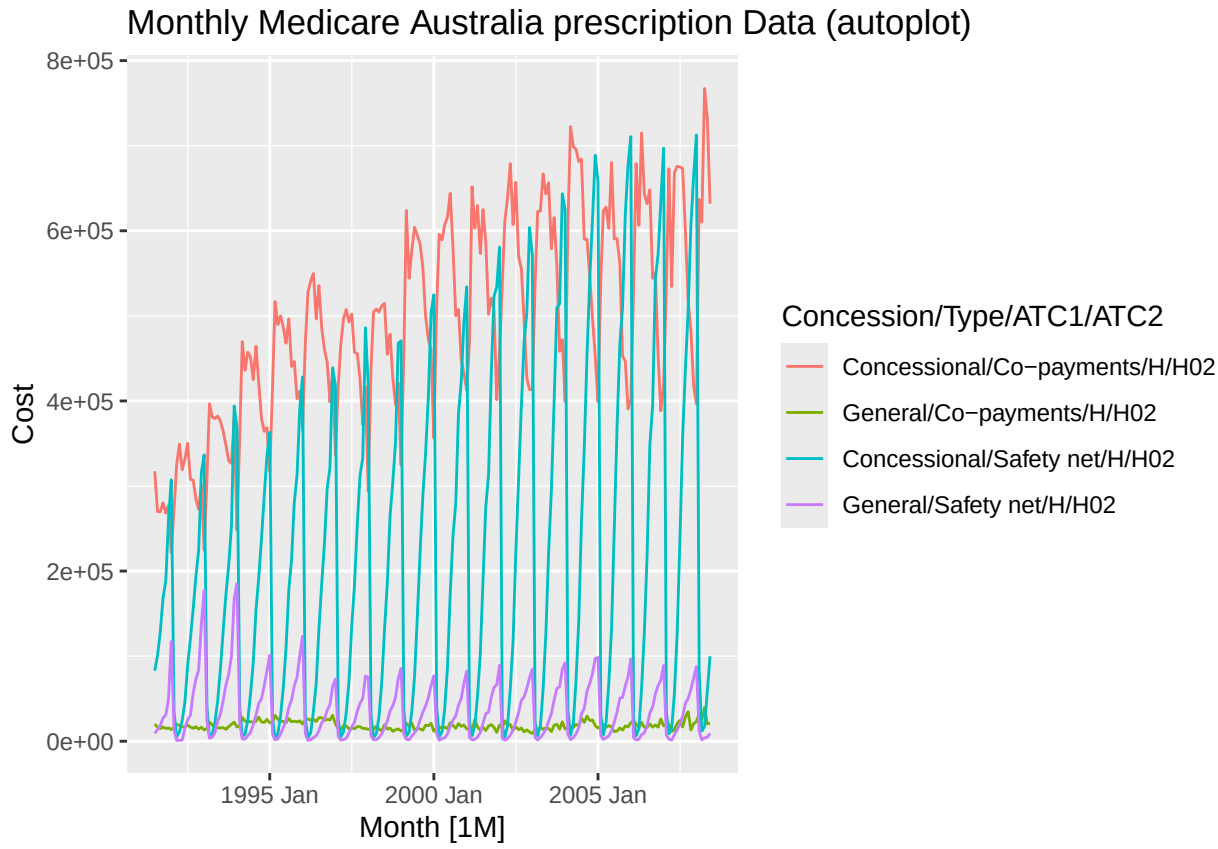
```
?PBS
```

```
PBS
```

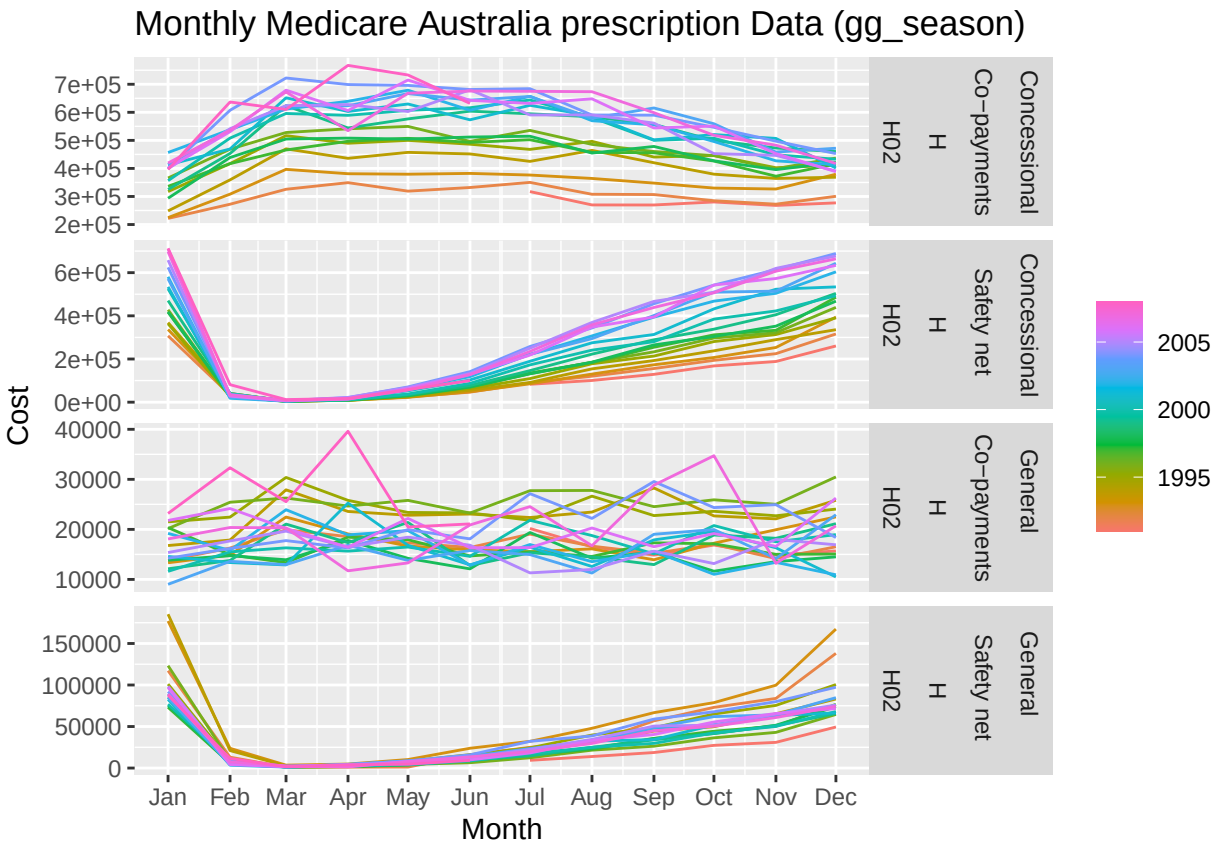
```
## # A tsibble: 67,596 x 9 [1M]
## # Key:       Concession, Type, ATC1, ATC2 [336]
##       Month Concession   Type      ATC1  ATC1_desc ATC2  ATC2_desc Scripts  Cost
##       <mth> <chr>        <chr>     <chr> <chr>     <chr> <chr>       <dbl> <dbl>
##  1 1991 Jul Concessional Co-payme~ A     Alimenta~ A01   STOMATOL~   18228 67877
##  2 1991 Aug Concessional Co-payme~ A     Alimenta~ A01   STOMATOL~   15327 57011
##  3 1991 Sep Concessional Co-payme~ A     Alimenta~ A01   STOMATOL~   14775 55020
##  4 1991 Oct Concessional Co-payme~ A     Alimenta~ A01   STOMATOL~   15380 57222
##  5 1991 Nov Concessional Co-payme~ A     Alimenta~ A01   STOMATOL~   14371 52120
##  6 1991 Dec Concessional Co-payme~ A     Alimenta~ A01   STOMATOL~   15028 54299
```

```
##  7 1992 Jan Concessional Co-payme~ A     Alimenta~ A01   STOMATOL~   11040 39753
##  8 1992 Feb Concessional Co-payme~ A     Alimenta~ A01   STOMATOL~   15165 54405
##  9 1992 Mar Concessional Co-payme~ A     Alimenta~ A01   STOMATOL~   16898 61108
## 10 1992 Apr Concessional Co-payme~ A     Alimenta~ A01   STOMATOL~   18141 65356
## # i 67,586 more rows
```

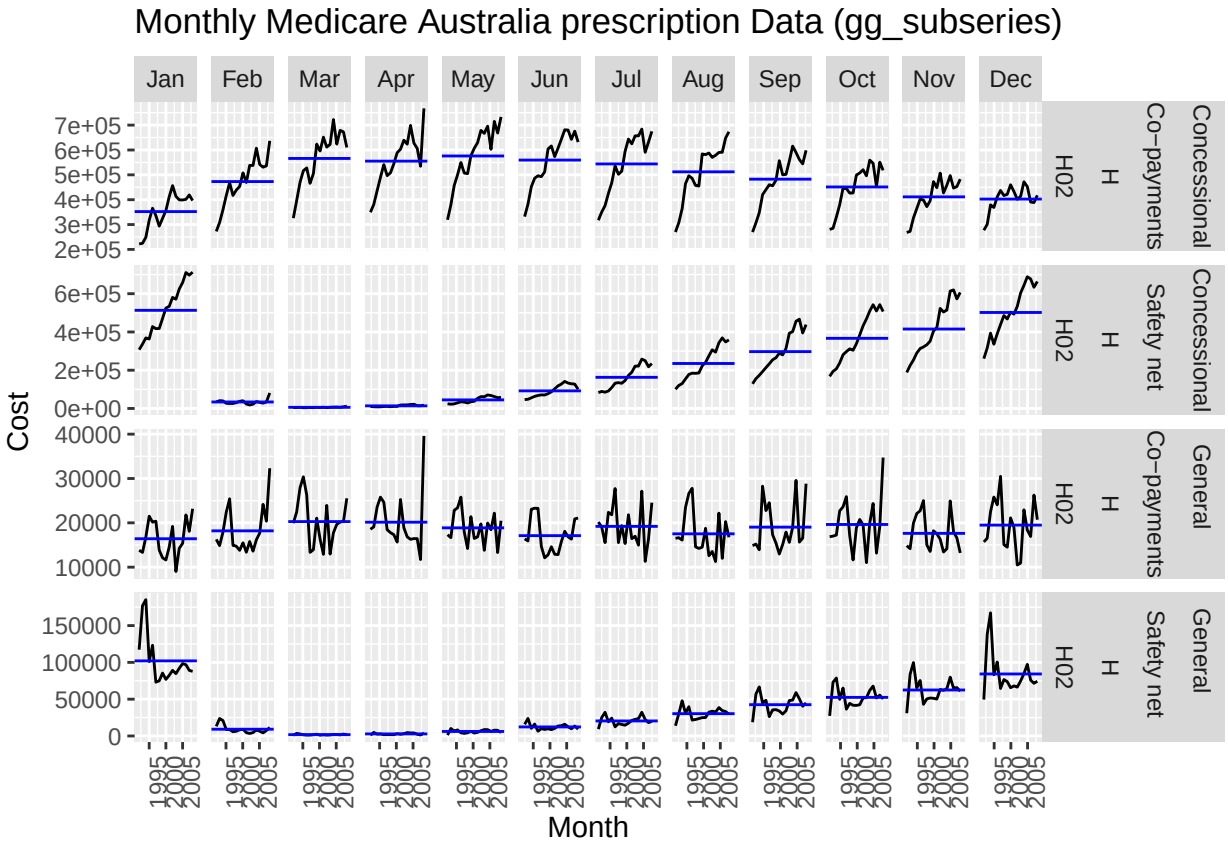
```
PBS %>%
  filter(ATC2 == "H02") %>%
autoplot(Cost)+
  labs(title = "Monthly Medicare Australia prescription Data (autoplot)")
```




```
PBS %>%
  filter(ATC2 == "H02") %>%
gg_season(Cost)+
  labs(title = "Monthly Medicare Australia prescription Data (gg_season)")
```

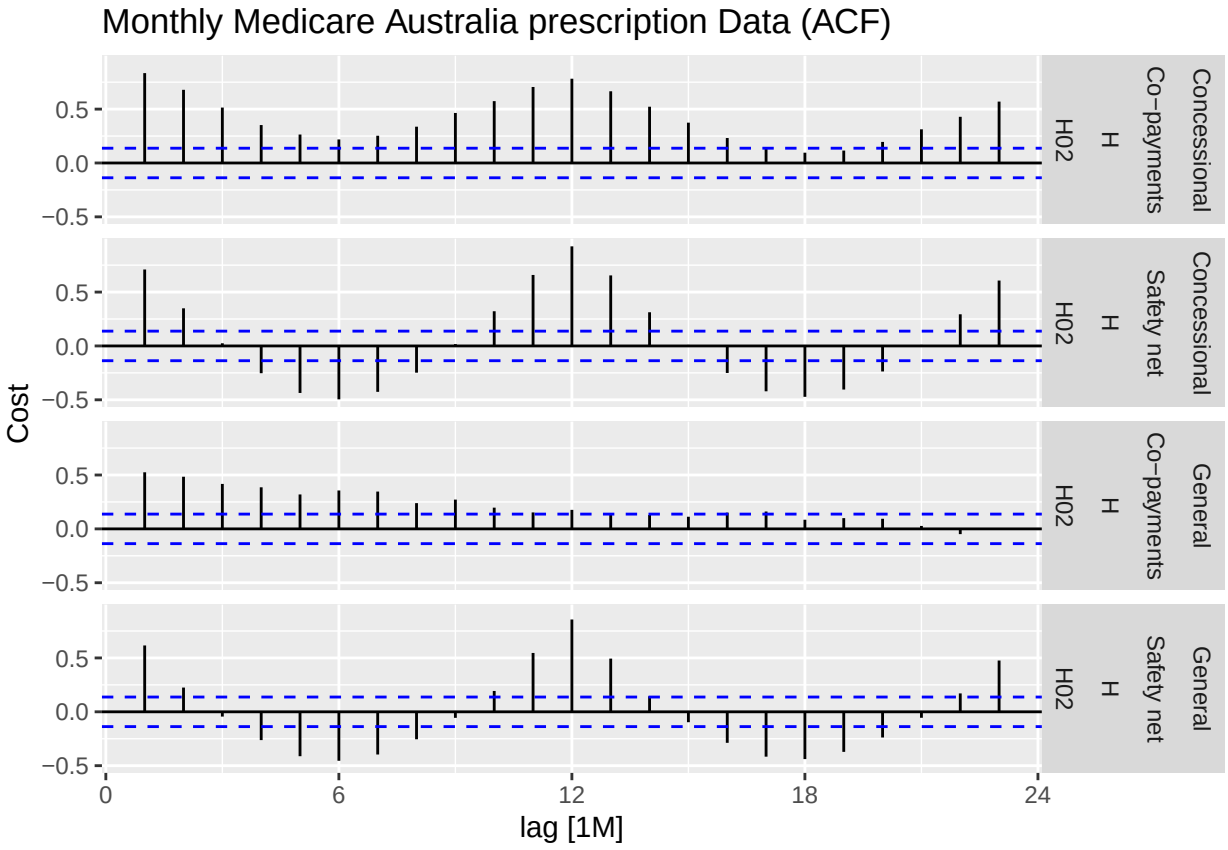



```
PBS %>%
  filter(ATC2 == "H02") %>%
gg_subseries(Cost)+
  labs(title = "Monthly Medicare Australia prescription Data (gg_subseries)")
```




```
#Fail to plot the gg_lag
#PBS %>%
  #filter(ATC2 == "H02") %>%
#gg_lag(Cost, geom = "point")+
  #labs(title = "Monthly Medicare Australia prescription Data (gg_lag)")
```

```
PBS %>%
  filter(ATC2 == "H02") %>%
  ACF(Cost)%>%
  autoplot()+
  labs(y = "Cost", title = "Monthly Medicare Australia prescription Data (ACF)")
```



us_gasoline Data:

- a.Can you spot any seasonality, cyclicity and trend?
- answer a: No seasonality, and cyclicity. But the main trend is going upward.
- b.What do you learn about the series?
- answer b:The series shows the upward trend.
- c.What can you say about the seasonal patterns?
- answer c: No seasonal patterns.
- d.Can you identify any unusual years?
- answer d:Around year 2017, the trend start to drop.

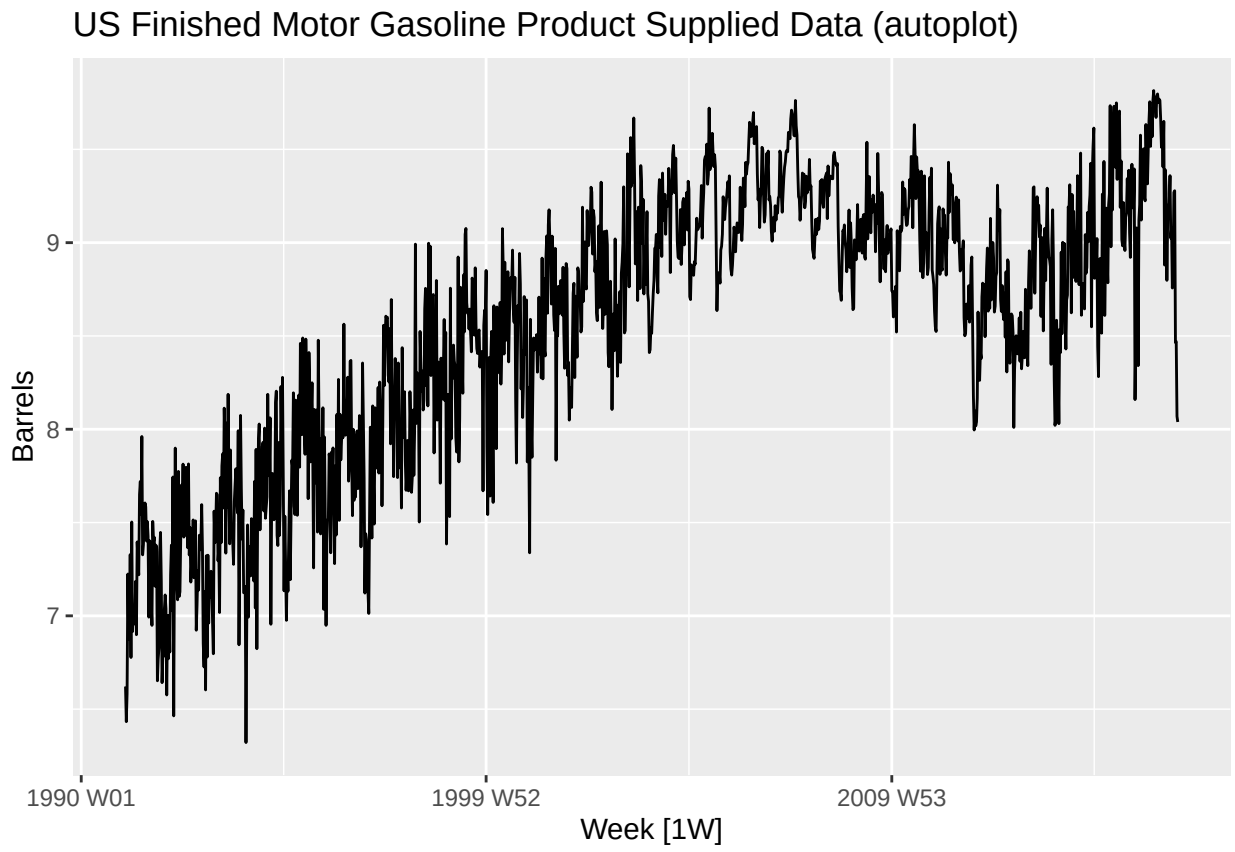
```
?us_gasoline
```

```
us_gasoline
```

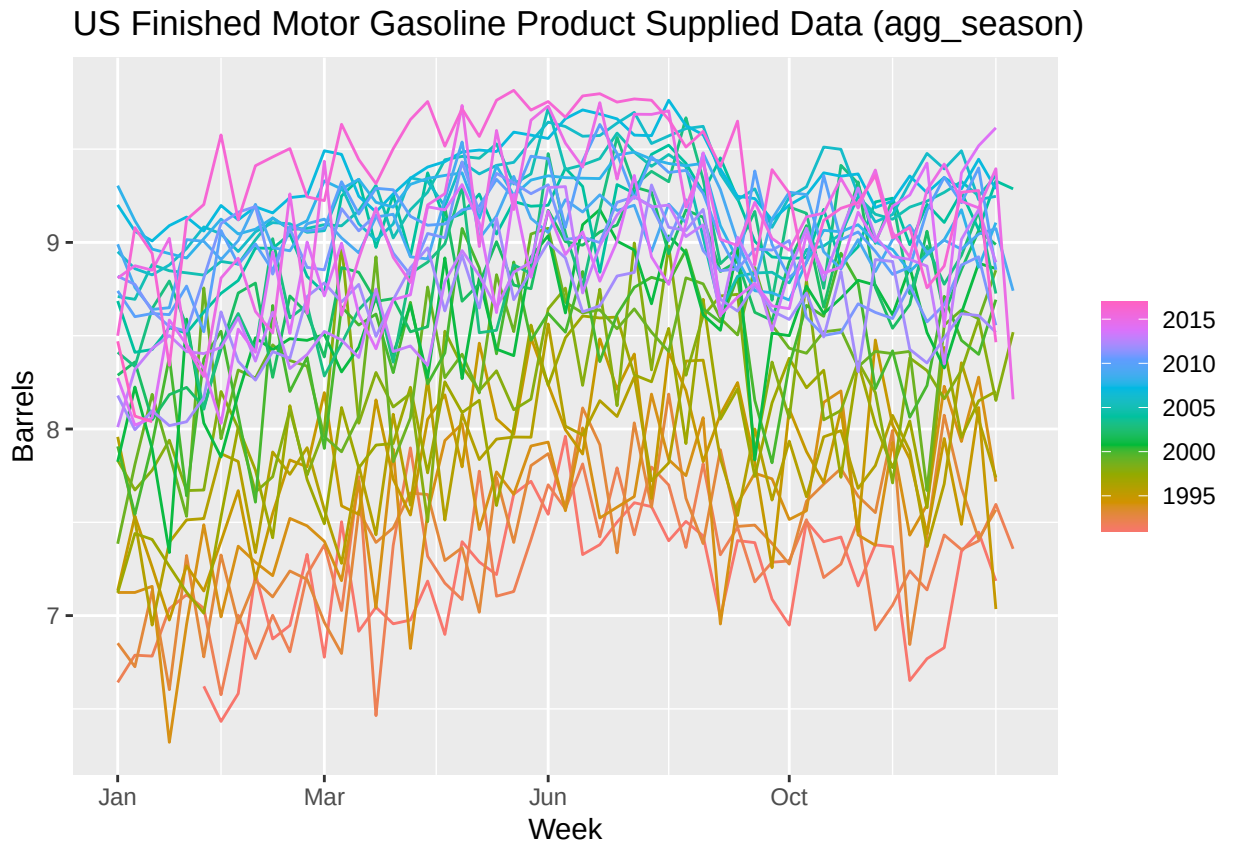
```
## # A tsibble: 1,355 x 2 [1W]
##        Week Barrels
##       <week>   <dbl>
##  1 1991 W06    6.62
##  2 1991 W07    6.43
##  3 1991 W08    6.58
##  4 1991 W09    7.22
##  5 1991 W10    6.88
##  6 1991 W11    6.95
```

```
##  7 1991 W12    7.33
##  8 1991 W13    6.78
##  9 1991 W14    7.50
## 10 1991 W15    6.92
## # i 1,345 more rows
```

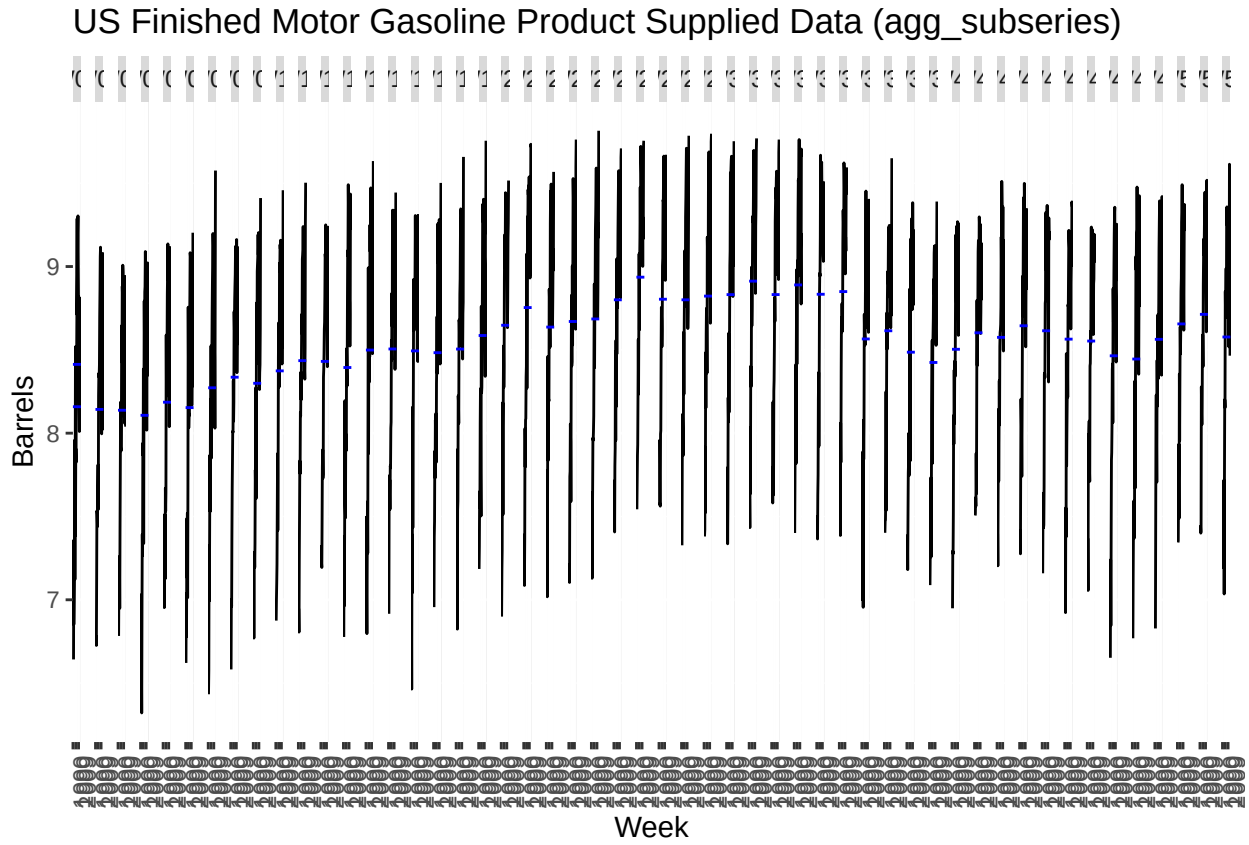
```
us_gasoline %>%
  autoplot(Barrels)+
  labs(title = "US Finished Motor Gasoline Product Supplied Data (autoplot)")
```



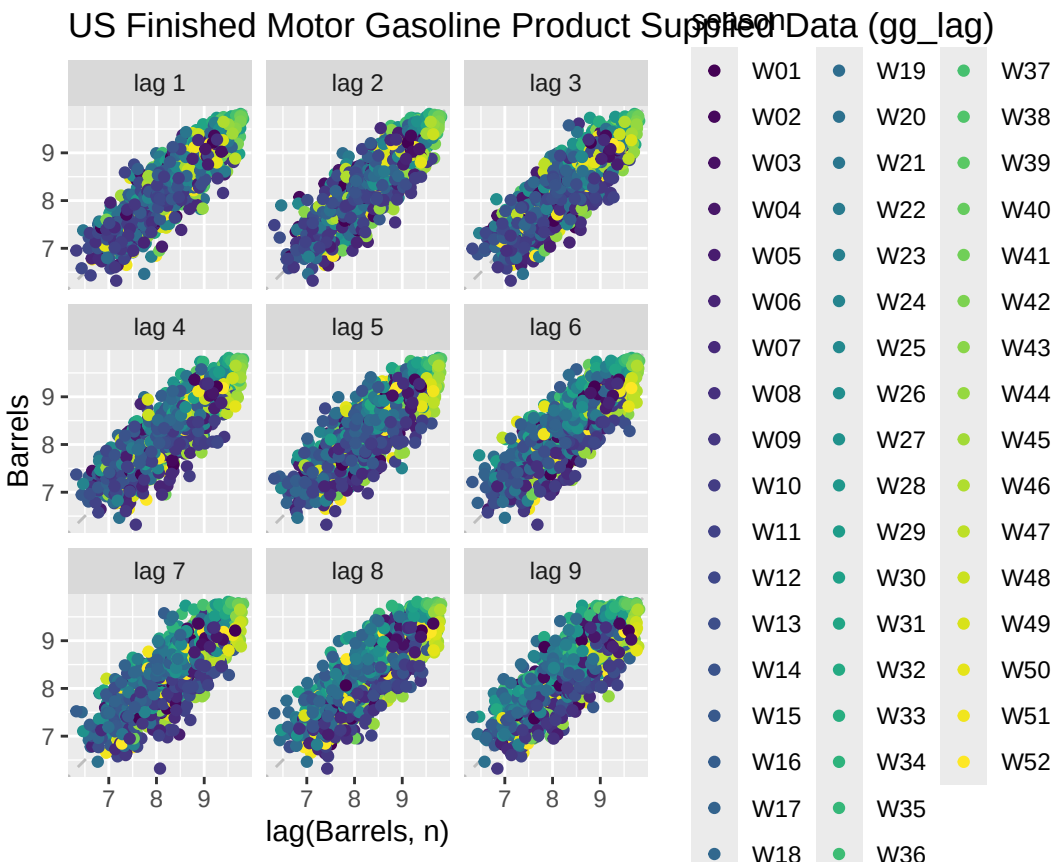

```
us_gasoline %>%
  gg_season(Barrels)+
  labs(title = "US Finished Motor Gasoline Product Supplied Data (agg_season)")
```



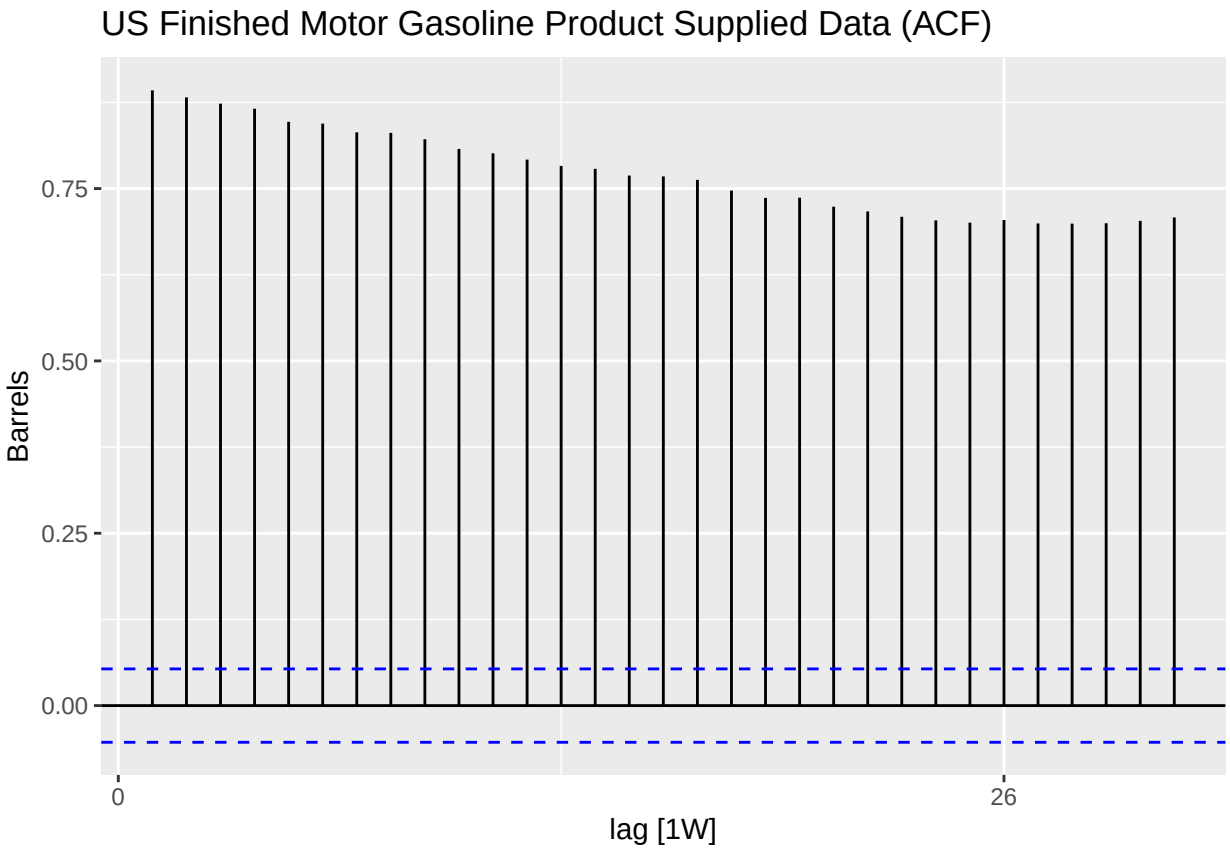

```
us_gasoline %>%
  gg_subseries(Barrels)+
  labs(title = "US Finished Motor Gasoline Product Supplied Data (agg_subseries)")
```




```
us_gasoline %>%
  gg_lag(Barrels, geom = "point")+
  labs(title = "US Finished Motor Gasoline Product Supplied Data (gg_lag)")
```




```
us_gasoline %>%
  ACF(Barrels)%>%
  autoplot()+
  labs(y = "Barrels", title = "US Finished Motor Gasoline Product Supplied Data (ACF)")
```


US Finished Motor Gasoline Product Supplied Data (ACF)
Barrels
0.75
0.50
0.25
0.00
0
26
lag [1W]