DATA 624: PREDICTIVE ANALYTICS

Gabriel Campos

Last edited February 04, 2024

```
library('fpp3')
library('tsibble')
library('ggplot2')
library('USgas')
library('readr')
library('zoo')
```

INSTRUCTIONS

Please submit exercises 2.1, 2.2, 2.3, 2.4, 2.5 and 2.8 from the Hyndman online Forecasting book. Please submit both your Rpubs link as well as attach the .pdf file with your code.

2.1

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

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

```
data("aus_production")
data("pelt")
data("gafa_stock")
data("vic_elec")
```

Bricks

i

Details Quarterly estimates of selected indicators of manufacturing production in Australia.

Bricks: Clay brick production in millions of bricks.

ii

Quarterly

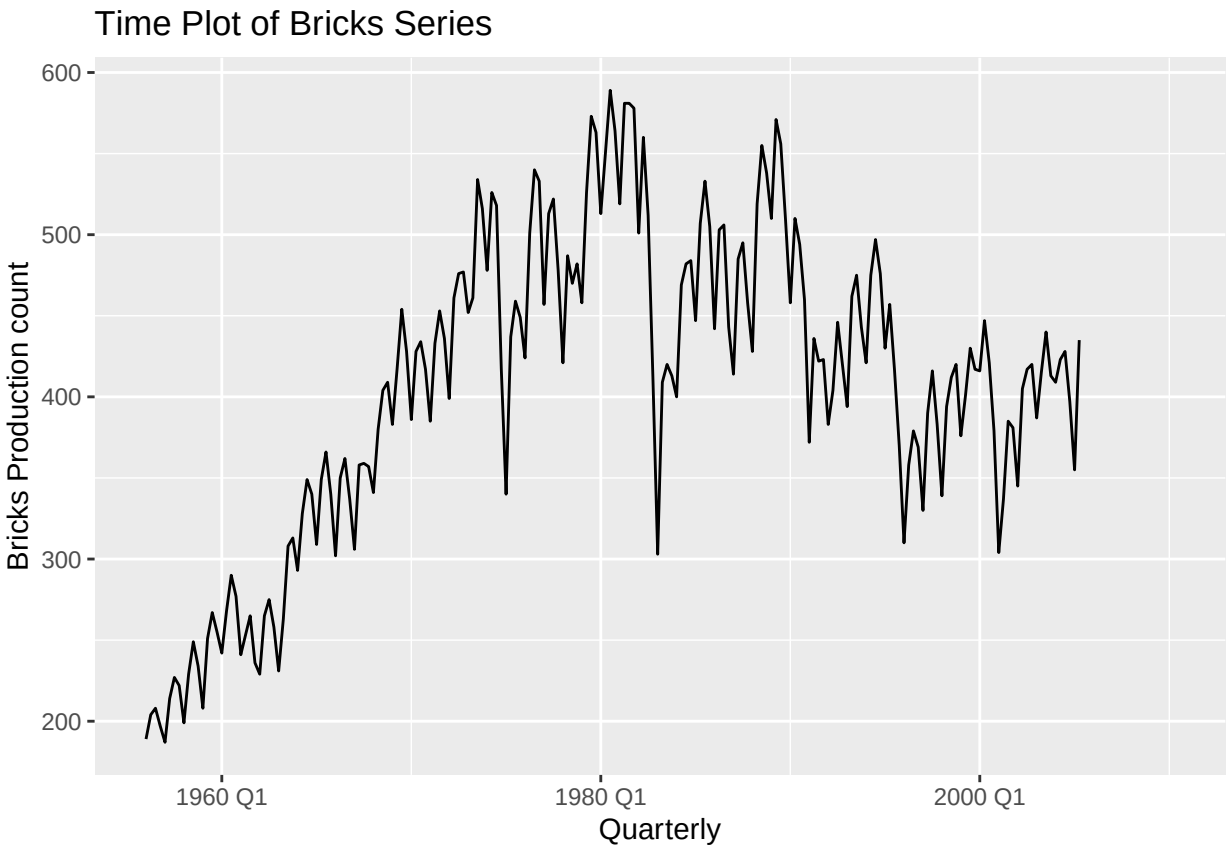
```
aus_production%>%
  select(Bricks)
```

```
## # A tsibble: 218 x 2 [1Q]
##    Bricks Quarter
##     <dbl>   <qtr>
##  1    189 1956 Q1
##  2    204 1956 Q2
##  3    208 1956 Q3
##  4    197 1956 Q4
##  5    187 1957 Q1
##  6    214 1957 Q2
##  7    227 1957 Q3
##  8    222 1957 Q4
##  9    199 1958 Q1
## 10    229 1958 Q2
## # i 208 more rows
```

iii

```
autoplot(aus_production,Bricks) +
  labs(title = "Time Plot of Bricks Series",
       x = "Quarterly",
       y = "Bricks Production count")
```

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



Lynx

i

pelt is an annual tsibble with two values:

Hare: The number of Snowshoe Hare pelts traded. Lynx: The number of Canadian Lynx pelts traded.

ii

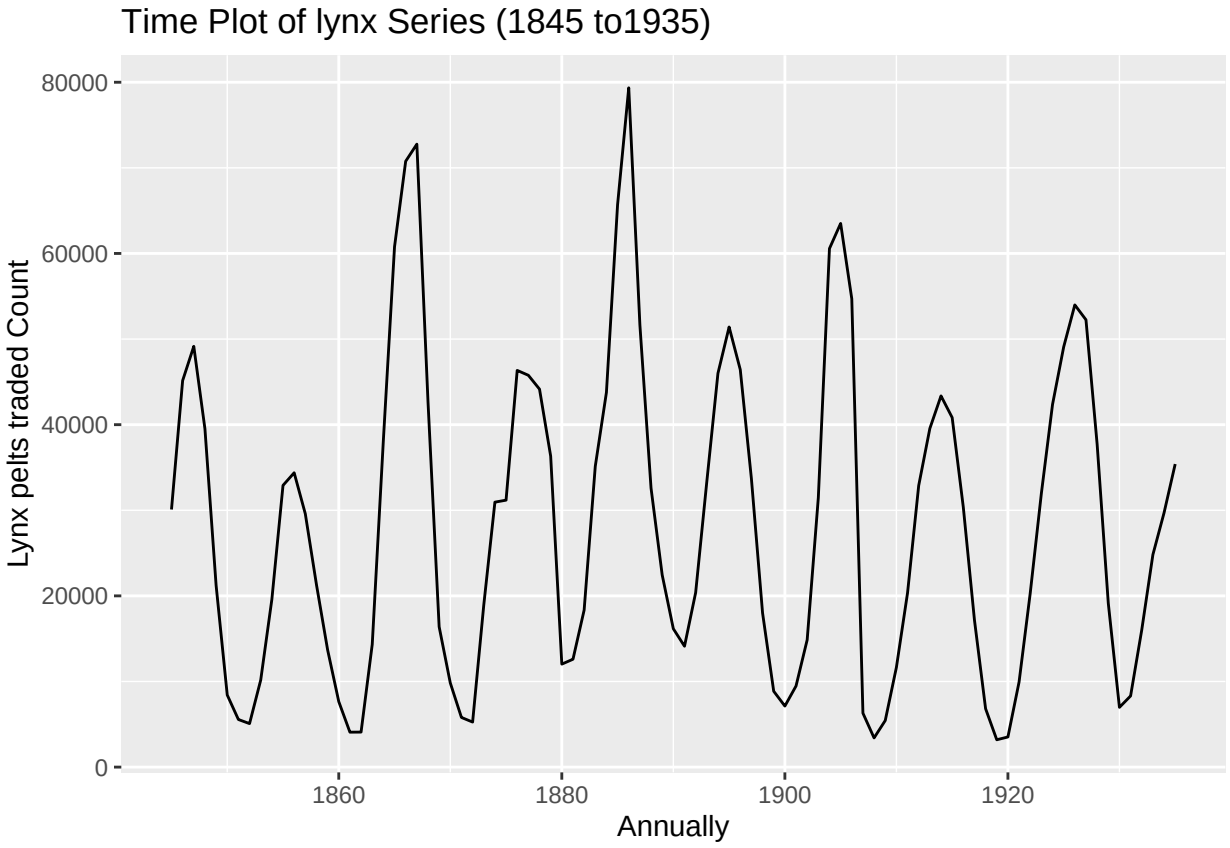
```
pelt %>%
  select(Lynx)
```

```
## # A tsibble: 91 x 2 [1Y]
##     Lynx  Year
##    <dbl> <dbl>
##  1 30090  1845
##  2 45150  1846
##  3 49150  1847
##  4 39520  1848
##  5 21230  1849
##  6  8420  1850
##  7  5560  1851
```

```
##  8  5080  1852
##  9 10170  1853
## 10 19600  1854
## # i 81 more rows
```

iii

```
autoplot(pelt,Lynx) +
  labs(title = "Time Plot of lynx Series (1845 to1935)",
       x = "Annually",
       y = "Lynx pelts traded Count")
```



Close

i

Details gafa_stock is a tsibble containing data on irregular trading days:

Open: The opening price for the stock. High: The stock's highest trading price. Low: The stock's lowest trading price. Close: The closing price for the stock. Adj_Close: The adjusted closing price for the stock. Volume: The amount of stock traded. Each stock is uniquely identified by one key:

Symbol: The ticker symbol for the stock.

ii

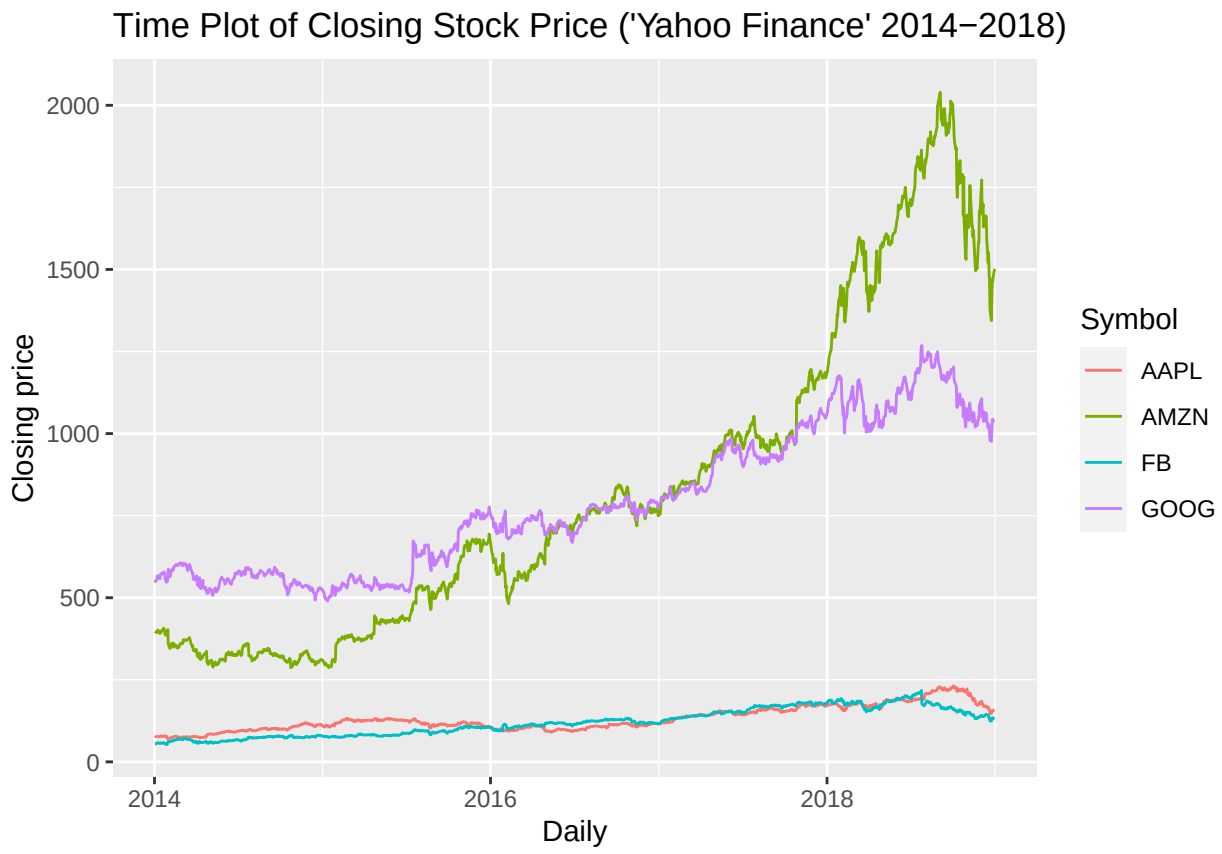
```
gafa_stock%>%
  select(Close)
```

```
## # A tsibble: 5,032 x 3 [!]
## # Key:       Symbol [4]
##    Close Date       Symbol
##    <dbl> <date>     <chr>
##  1  79.0 2014-01-02 AAPL
##  2  77.3 2014-01-03 AAPL
##  3  77.7 2014-01-06 AAPL
##  4  77.1 2014-01-07 AAPL
##  5  77.6 2014-01-08 AAPL
##  6  76.6 2014-01-09 AAPL
##  7  76.1 2014-01-10 AAPL
##  8  76.5 2014-01-13 AAPL
##  9  78.1 2014-01-14 AAPL
## 10  79.6 2014-01-15 AAPL
## # i 5,022 more rows
```

The `gafa_stock` is daily data

iii

```
autoplot(gafa_stock,Close) +
  labs(title = "Time Plot of Closing Stock Price ('Yahoo Finance' 2014-2018)",
       x = "Daily",
       y = "Closing price")
```



Demand

i

Description*

`vic_elec` is a half-hourly `tsibble` with three values:

Demand: Total electricity demand in MWh. Temperature: Temperature of Melbourne (BOM site 086071). Holiday: Indicator for if that day is a public holiday.

ii

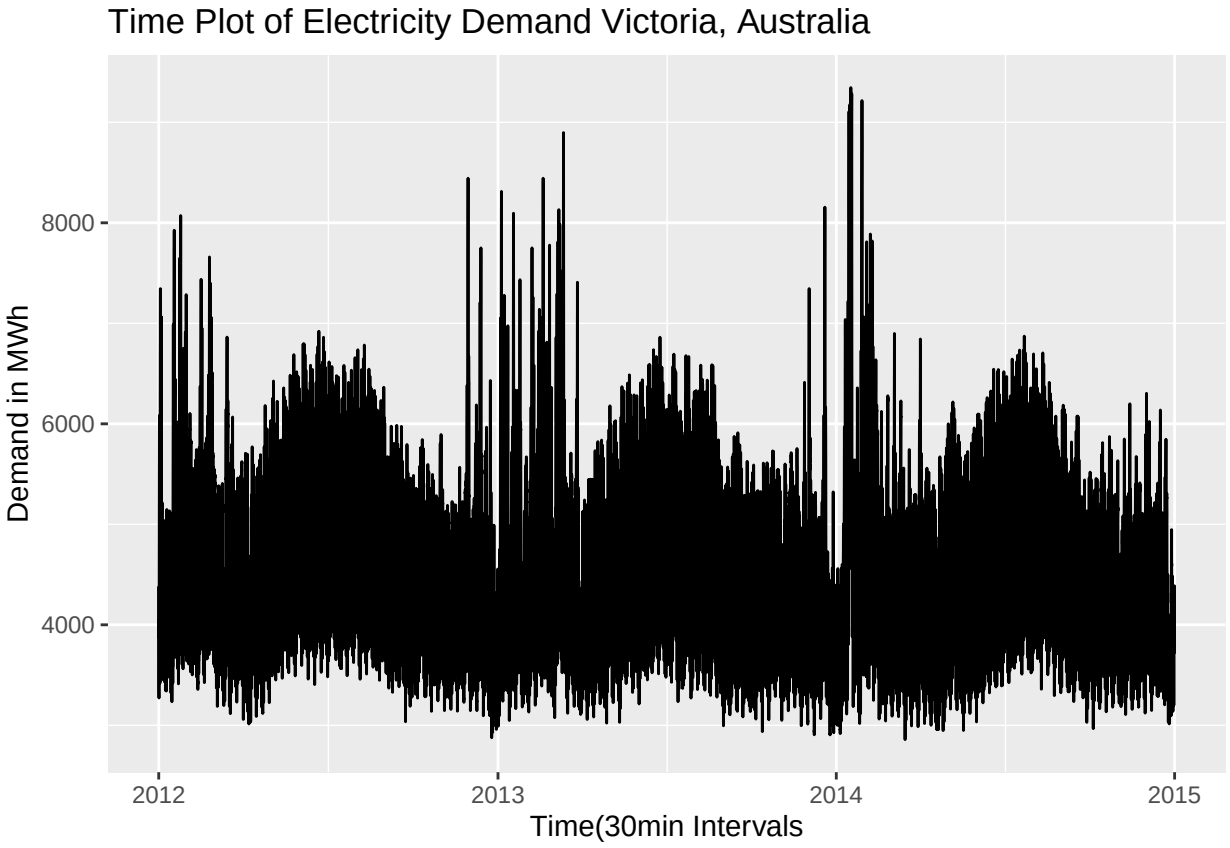
```
vic_elec %>%
  select(Demand)
```

```
## # A tsibble: 52,608 x 2 [30m] <Australia/Melbourne>
##    Demand Time
##     <dbl> <dttm>
##  1  4383. 2012-01-01 00:00:00
##  2  4263. 2012-01-01 00:30:00
##  3  4049. 2012-01-01 01:00:00
##  4  3878. 2012-01-01 01:30:00
##  5  4036. 2012-01-01 02:00:00
```

```
##  6  3866. 2012-01-01 02:30:00
##  7  3694. 2012-01-01 03:00:00
##  8  3562. 2012-01-01 03:30:00
##  9  3433. 2012-01-01 04:00:00
## 10  3359. 2012-01-01 04:30:00
## # i 52,598 more rows
```

iii & vi

```
autoplot(vic_elec,Demand) +
  labs(title = "Time Plot of Electricity Demand Victoria, Australia",
       x = "Time(30min Intervals",
       y = "Demand in MWh")
```



2.2

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

```
colnames(gafa_stock)
```

```
## [1] "Symbol"    "Date"      "Open"      "High"      "Low"       "Close"
## [7] "Adj_Close" "Volume"
```

```
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
```

2.3

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.

a.

You can read the data into R with the following script:

```
df_tute1 <- readr::read_csv(tute1_csv)
head(df_tute1,20)
```

```
## # A tibble: 20 x 4
##    Quarter    Sales AdBudget   GDP
##    <date>     <dbl>    <dbl> <dbl>
##  1 1981-03-01 1020.     659.  252.
##  2 1981-06-01  889.     589   291.
##  3 1981-09-01  795      512.  291.
##  4 1981-12-01 1004.     614.  292.
##  5 1982-03-01 1058.     647.  279.
##  6 1982-06-01  944.     602   254
##  7 1982-09-01  778.     531.  296.
##  8 1982-12-01  932.     608.  272.
##  9 1983-03-01  996.     638.  260.
## 10 1983-06-01  908.     582.  280.
## 11 1983-09-01  735.     507.  287.
## 12 1983-12-01  958.     607.  278
## 13 1984-03-01 1034.     659.  257.
## 14 1984-06-01  993.     615.  271
## 15 1984-09-01  792.     490.  301.
## 16 1984-12-01  914.     586.  290.
## 17 1985-03-01 1106.     663   267.
## 18 1985-06-01  985.     592.  274.
## 19 1985-09-01  824.     502.  301.
## 20 1985-12-01 1025.     616.  286.
```

b.

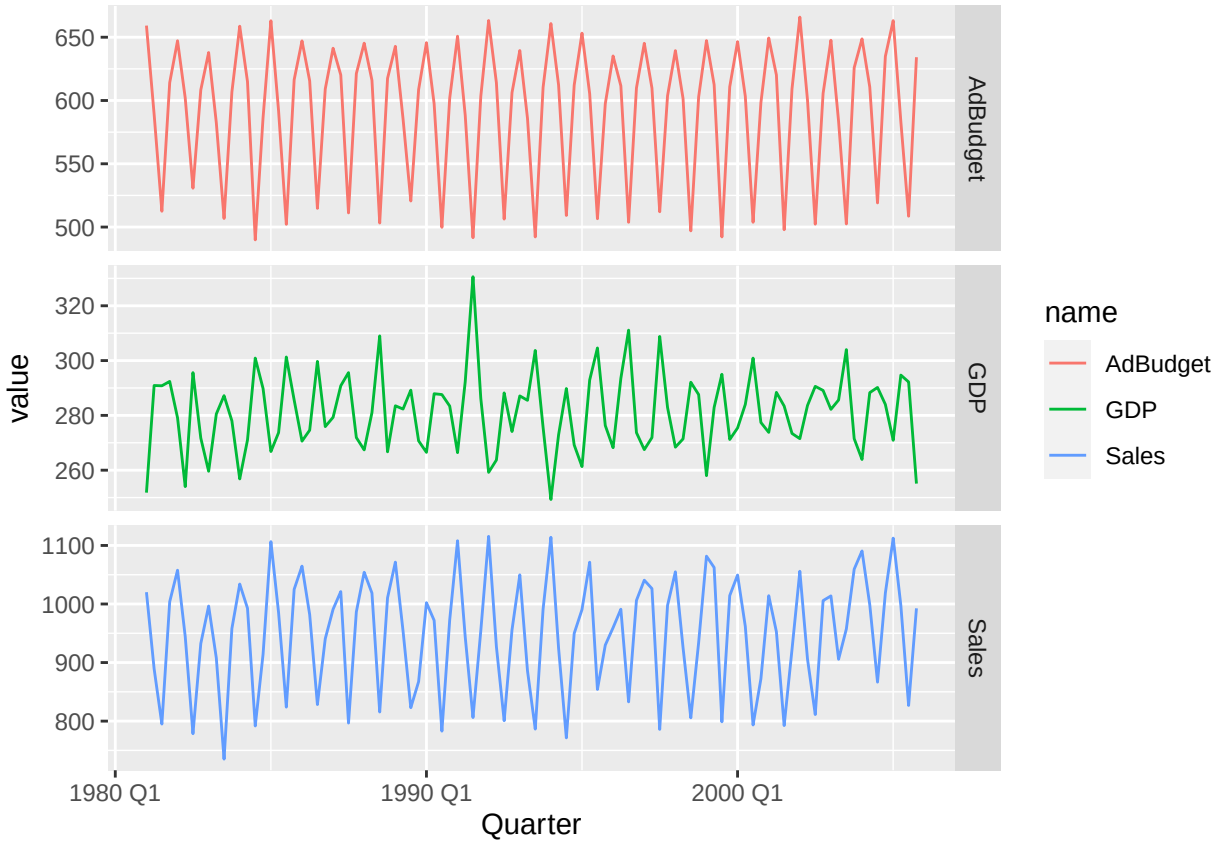
Convert the data to time series

```
mytimeseries <- df_tute1 |>
  mutate(Quarter = yearquarter(Quarter)) |>
  as_tsibble(index = Quarter)
```

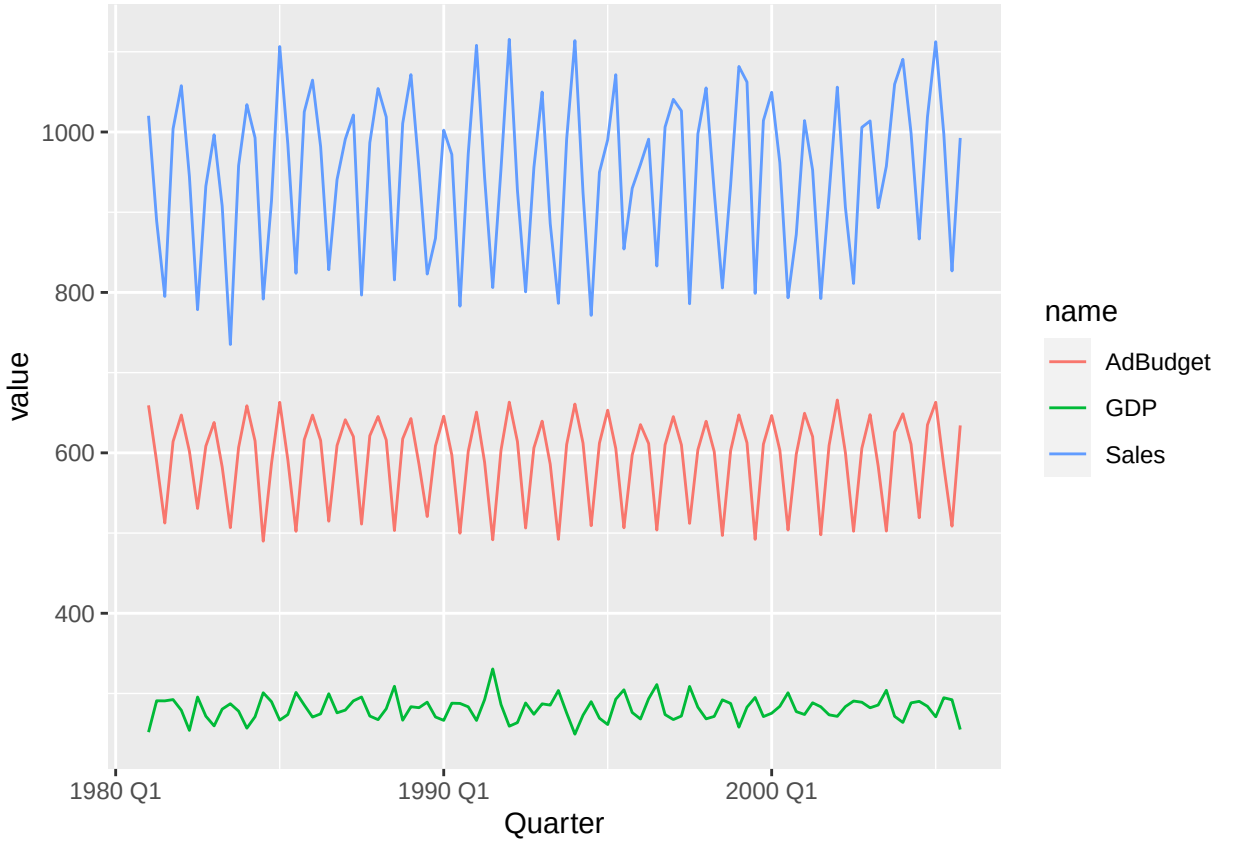
c.

Construct time series plots of each of the three series

```
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() #+
```




```
  # facet_grid(name ~ ., scales = "free_y")
```

The plot is encompassed in one plot without the `facet_grid()` function.

2.4

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

i. Install the `USgas` package.
ii. Create a `tsibble` from `us_total` with year as the index and state as the key.
iii. 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).

i

```
str(USgas::us_total)
```

```
## 'data.frame':    1266 obs. of  3 variables:
##  $ year : int  1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 ...
##  $ state: chr  "Alabama" "Alabama" "Alabama" "Alabama" ...
##  $ y    : int  324158 329134 337270 353614 332693 379343 350345 382367 353156 391093 ...
```

ii

Example

Forecasting Principles & Practice: 2.1 tsibble objects

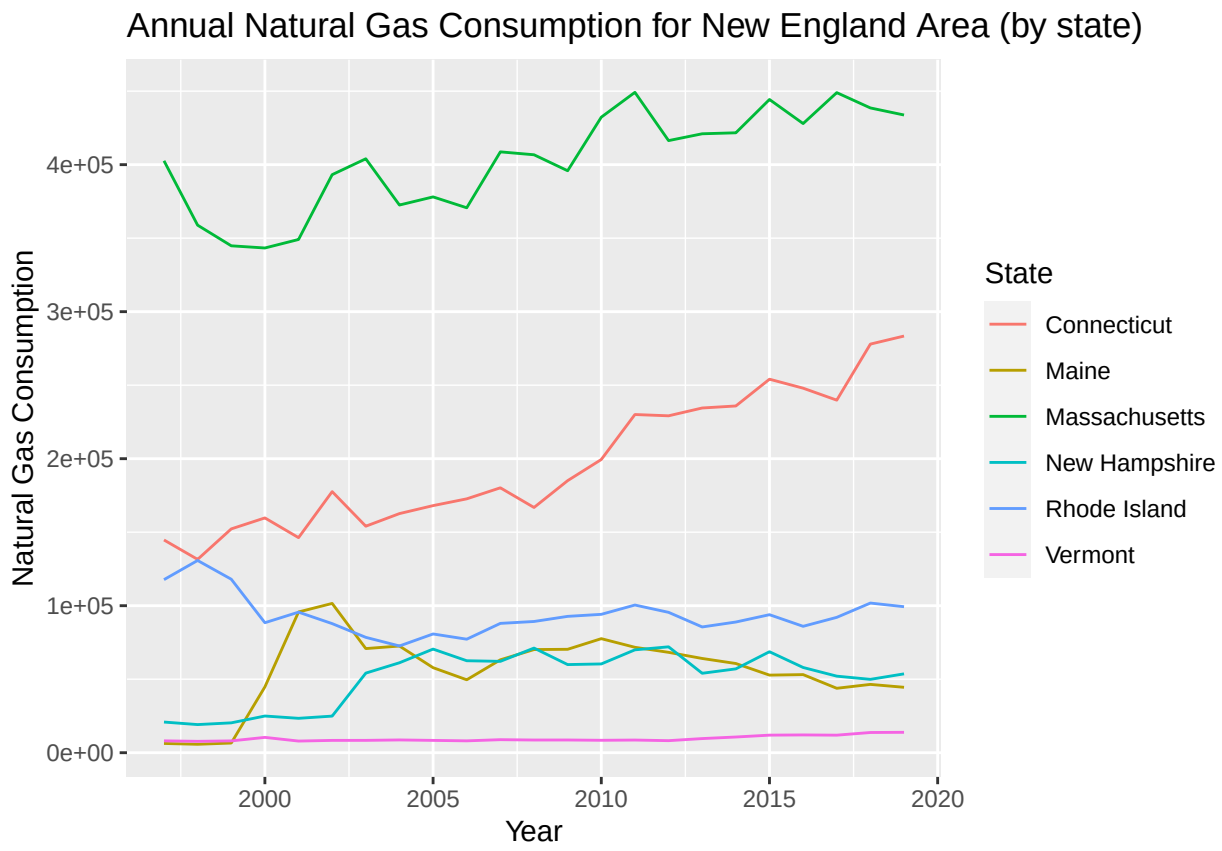
Template

```
    mydata <- tsibble(
        year = 2015:2019,
        y=c(123,39,78,52,110),
        index = year
    )
    mydata
```

```
mydata <- tsibble(
  state = us_total$state,
  year = us_total$year,
  value = us_total$y,
  index = year,
  key = state
)%>%
  filter(state %in% c("Maine", "Vermont", "New Hampshire", "Massachusetts", "Connecticut", "Rhode Islan
```

iii

```
ggplot(mydata, aes(x = year, y = value, color = state)) +
  geom_line() +
  labs(
    title = "Annual Natural Gas Consumption for New England Area (by state)",
    x = "Year",
    y = "Natural Gas Consumption",
    color = "State"
  )
```



2.5

a.

Download `tourism.xlsx` from the book website and read it into R using `readxl::read_excel()`.

```
PATH<-"C:/Users/Lenny/Documents/GitableGabe/Data624_Data/"
```

```
tourism_str <- paste(PATH,"tourism.xlsx", sep = "")
df_tourism <- readxl::read_excel(tourism_str)
rm(tourism_str)
tourism
```

```
## # A tsibble: 24,320 x 5 [1Q]
## # Key:       Region, State, Purpose [304]
##    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.
```

```
##  7 1999 Q3 Adelaide South Australia Business  169.
##  8 1999 Q4 Adelaide South Australia Business  134.
##  9 2000 Q1 Adelaide South Australia Business  154.
## 10 2000 Q2 Adelaide South Australia Business  169.
## # i 24,310 more rows
```

b.

Create a `tsibble` which is identical to the tourism `tsibble` from the `tsibble` package.

```
str(df_tourism)
```

```
## tibble [24,320 x 5] (S3: tbl_df/tbl/data.frame)
##  $ Quarter: chr [1:24320] "1998-01-01" "1998-04-01" "1998-07-01" "1998-10-01" ...
##  $ Region : chr [1:24320] "Adelaide" "Adelaide" "Adelaide" "Adelaide" ...
##  $ State  : chr [1:24320] "South Australia" "South Australia" "South Australia" "South Australia" ..
##  $ Purpose: chr [1:24320] "Business" "Business" "Business" "Business" ...
##  $ Trips  : num [1:24320] 135 110 166 127 137 ...
```

Example

Forecasting Principles & Practice: 2.1 tsibble objects

Template

```
  prison<- read::read_csv("data/prison_population.csv") %>%
            mutate(Quarter = yearquarter(date)) %>%
              select(-date) %>%
              as_tsibble(
              index = Quarter,
              key=c(state,gender,legal,indigenous)
              )
```

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

tibble_tourism
```

```
## # A tsibble: 24,320 x 5 [1Q]
## # Key:       Region, State, Purpose [304]
##    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.
##  7 1999 Q3 Adelaide South Australia Business  169.
##  8 1999 Q4 Adelaide South Australia Business  134.
##  9 2000 Q1 Adelaide South Australia Business  154.
## 10 2000 Q2 Adelaide South Australia Business  169.
## # i 24,310 more rows
```

c.

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

```
tibble_tourism %>%
  group_by(Region,Purpose)%>%
  summarize(TripsAvg = mean(Trips))%>%
  filter(TripsAvg == max(TripsAvg))%>%
  arrange(desc(TripsAvg))
```

```
## # A tsibble: 76 x 4 [1Q]
## # Key:       Region, Purpose [76]
## # Groups:    Region [76]
##    Region                 Purpose  Quarter TripsAvg
##    <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
```

d.

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

```
tibble_tourism_v2 <- tibble_tourism %>%
  group_by(State)%>%
  summarize(Total=sum(Trips))

tibble_tourism_v2
```

```
## # A tsibble: 640 x 3 [1Q]
## # Key:       State [8]
##    State Quarter Total
##    <chr>   <qtr> <dbl>
##  1 ACT   1998 Q1  551.
##  2 ACT   1998 Q2  416.
##  3 ACT   1998 Q3  436.
##  4 ACT   1998 Q4  450.
##  5 ACT   1999 Q1  379.
##  6 ACT   1999 Q2  558.
##  7 ACT   1999 Q3  449.
##  8 ACT   1999 Q4  595.
##  9 ACT   2000 Q1  600.
## 10 ACT   2000 Q2  557.
## # i 630 more rows
```

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`.

i. Can you spot any seasonality, cyclicity and trend?
ii. What do you learn about the series?
iii. What can you say about the seasonal patterns?
iv. Can you identify any unusual years?

Total Private

Example

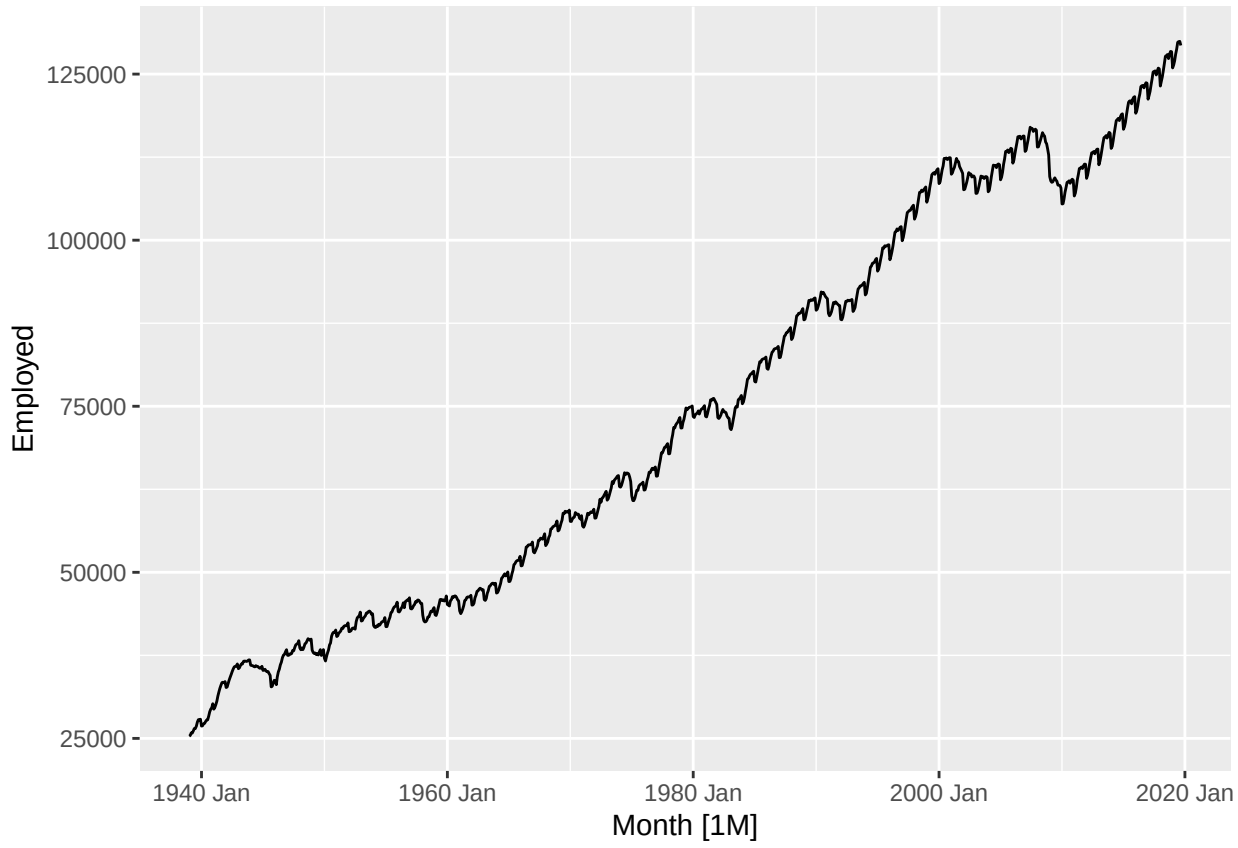
```
  vic_elec |> gg_season(Demand, period = "day") +
    theme(legend.position = "none") +
    labs(y="MWh", title="Electricity demand: Victoria")
```

```
us_employment
```

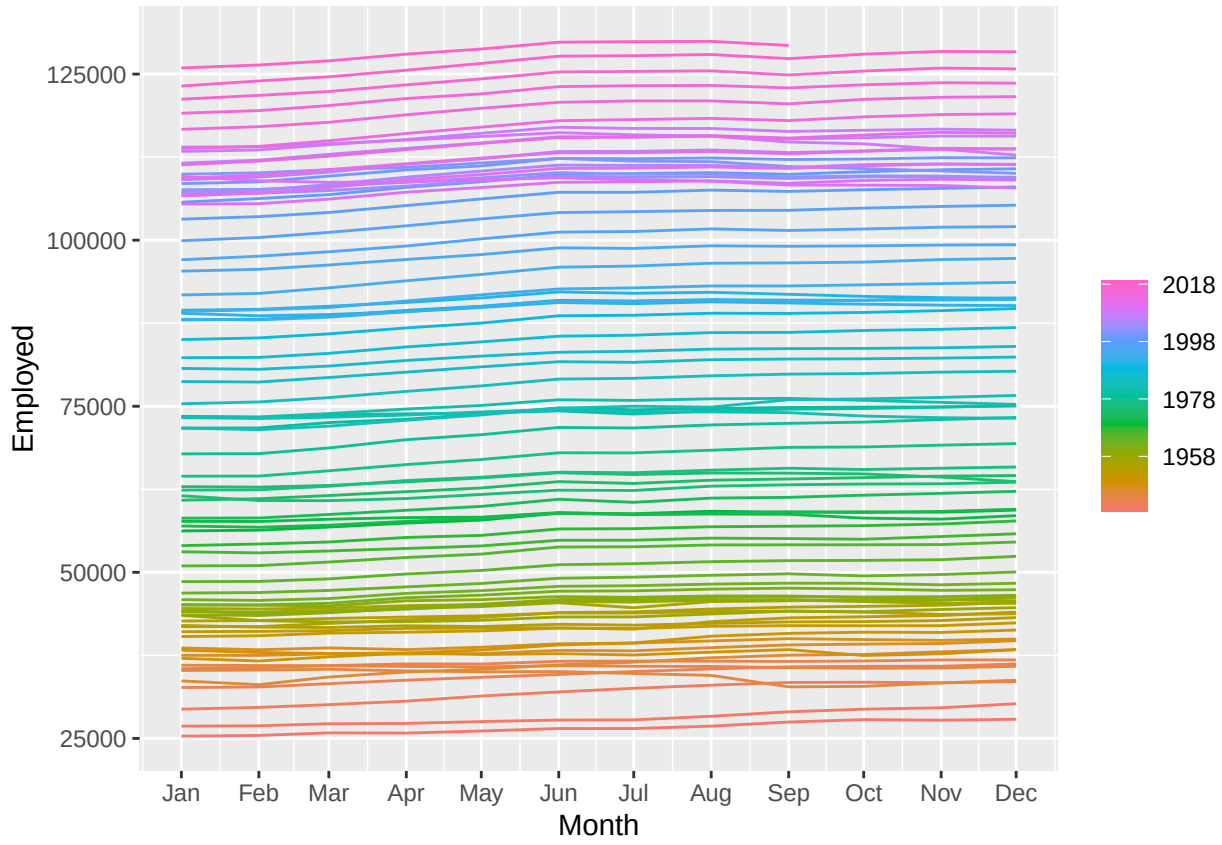
```
## # A tsibble: 143,412 x 4 [1M]
## # Key:       Series_ID [148]
##       Month Series_ID     Title          Employed
##       <mth> <chr>         <chr>             <dbl>
##  1 1939 Jan CEU0500000001 Total Private     25338
##  2 1939 Feb CEU0500000001 Total Private     25447
##  3 1939 Mar CEU0500000001 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,period="month")
```

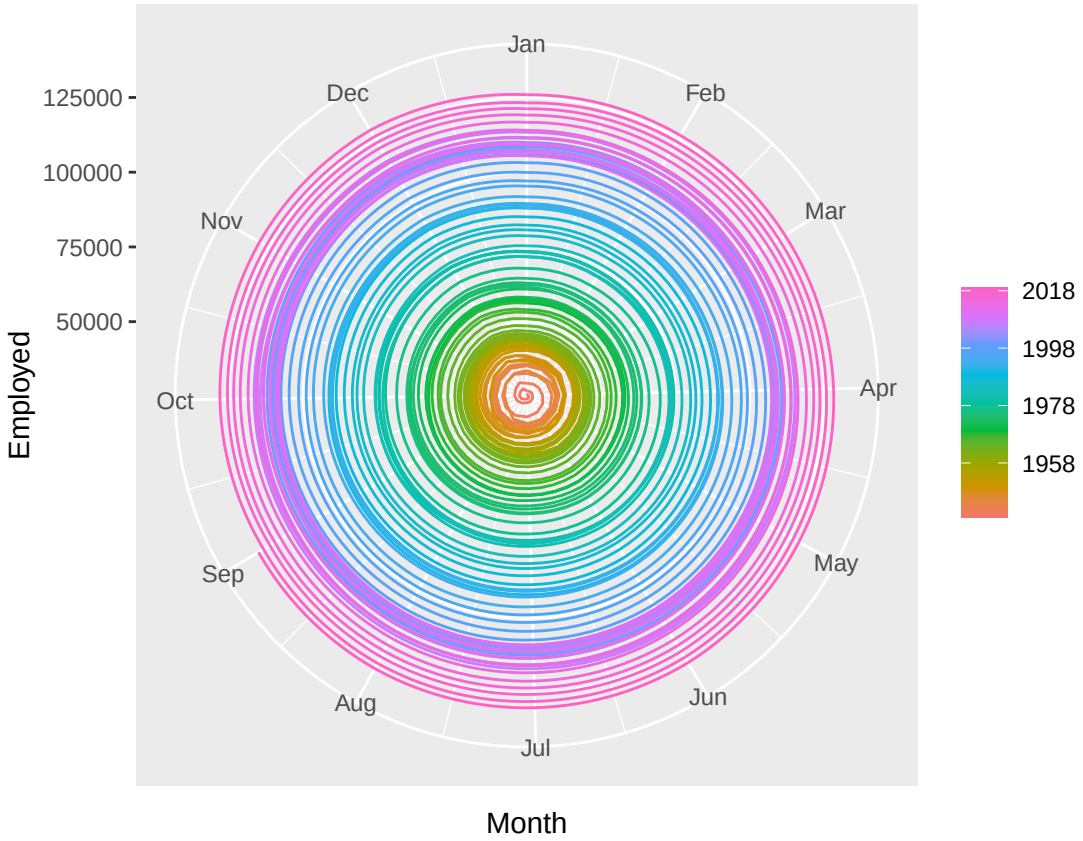
```
## Warning in geom_line(...): Ignoring unknown parameters: `period`
```



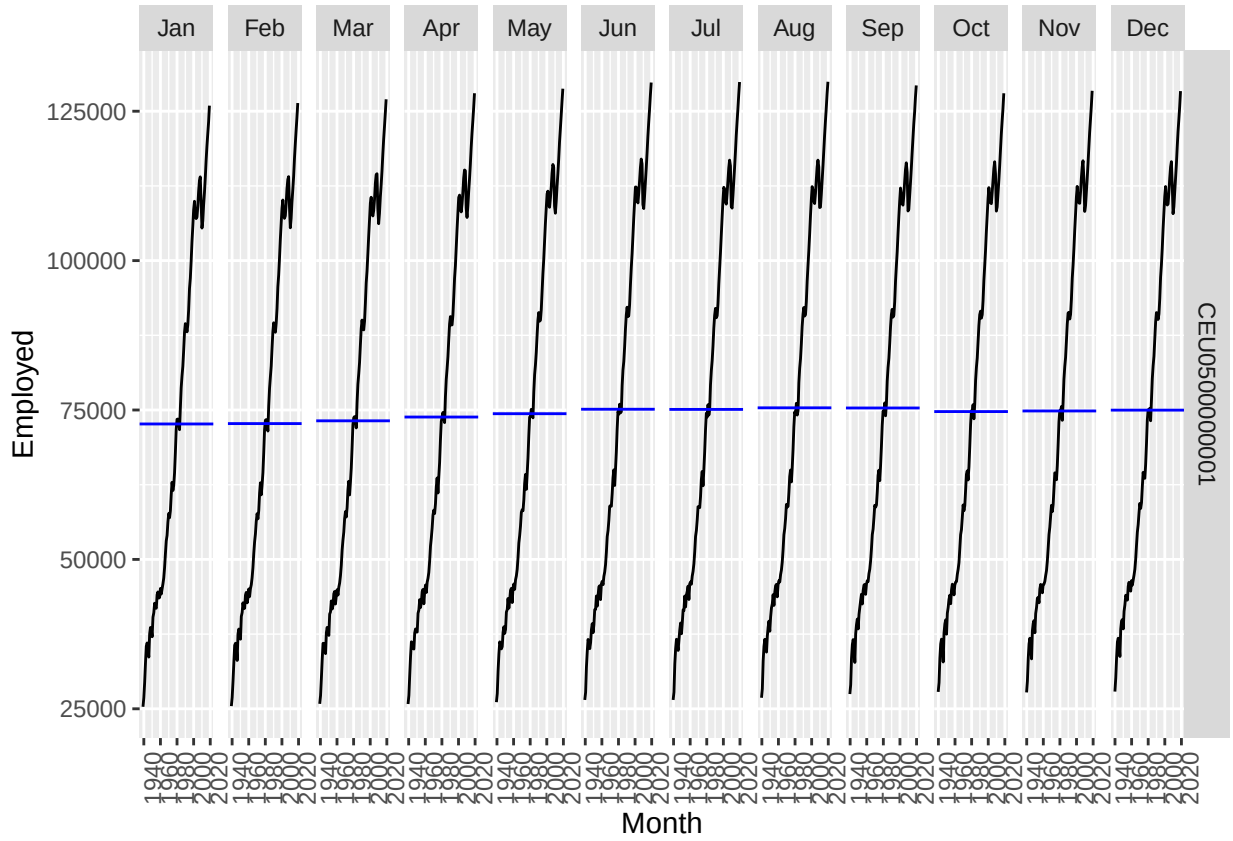

```
us_employment%>%
  filter(Title=="Total Private")%>%
  gg_season(Employed, polar = FALSE)
```



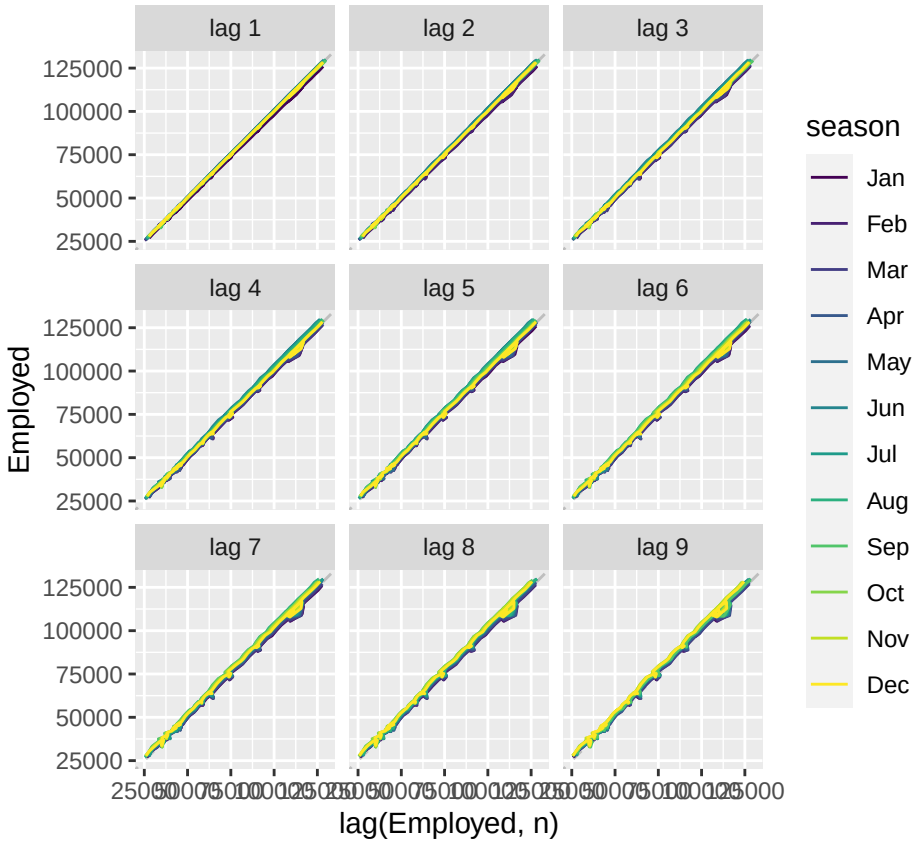

```
us_employment%>%
  filter(Title=="Total Private")%>%
  gg_season(Employed, polar = TRUE)
```



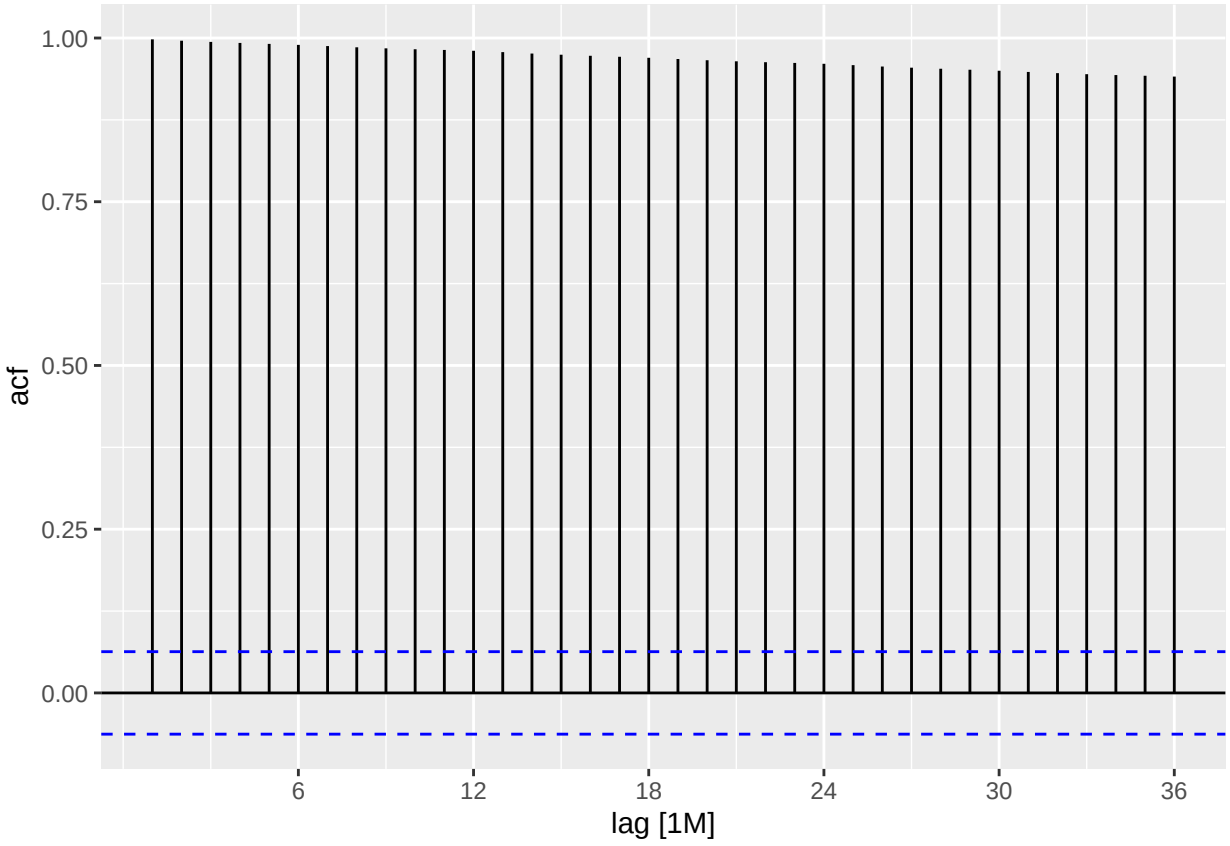

```
us_employment%>%
  filter(Title=="Total Private")%>%
  gg_subseries(Employed)
```




```
us_employment%>%
  filter(Title=="Total Private")%>%
  gg_lag(Employed)
```




```
us_employment%>%
  filter(Title=="Total Private")%>%
  ACF(us_employment$Employed)%>%
  autoplot()
```



i

There is a clear upwards trend in small increments for the data.

ii

Growth has been consistent without any extreme spike or drop.

iii

No Seasonality is noted indicating there is not particular season with an affect on employment positive or negative.

iv

A small dip around 2010 which I believe aligns with the recession.

Bricks

```
aus_production
```

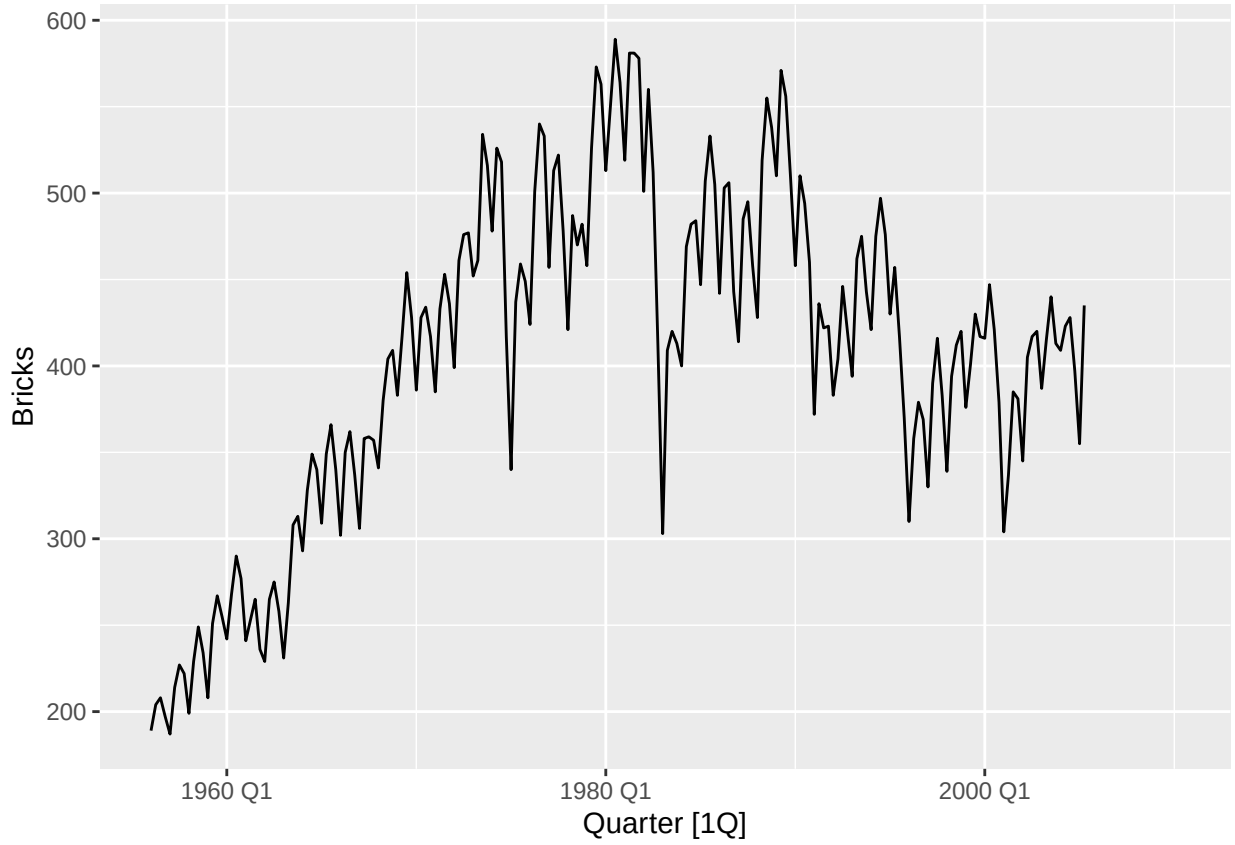
```
## # A tsibble: 218 x 7 [1Q]
##    Quarter  Beer Tobacco Bricks Cement Electricity   Gas
##      <qtr> <dbl>   <dbl>  <dbl>  <dbl>       <dbl> <dbl>
##  1 1956 Q1   284    5225    189    465        3923     5
##  2 1956 Q2   213    5178    204    532        4436     6
##  3 1956 Q3   227    5297    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%>%
  select(Bricks)%>%
  autoplot(period="quarter")
```

```
## Plot variable not specified, automatically selected `.vars = Bricks`
```

```
## Warning in geom_line(...): Ignoring unknown parameters: `period`
```

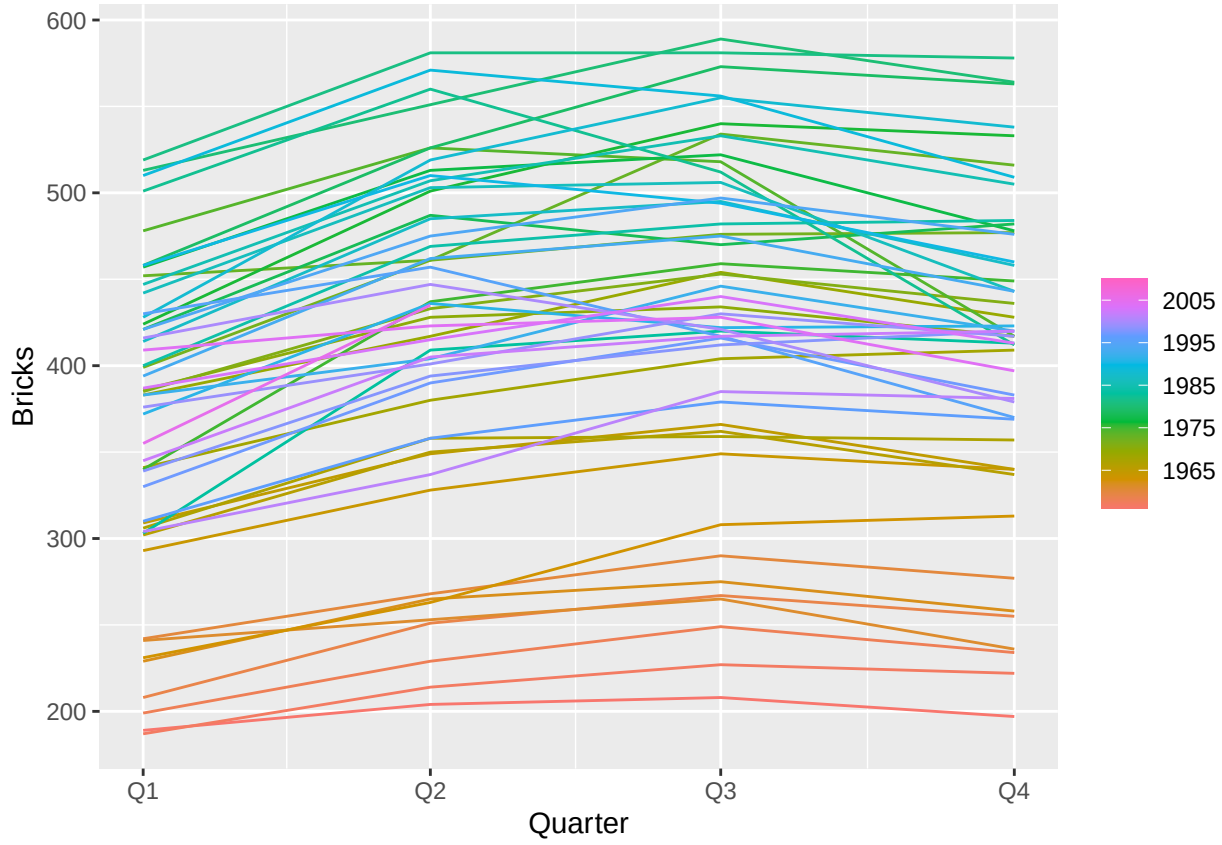
```
## Warning: Removed 20 rows containing missing values (`geom_line()`).
```



```
aus_production%>%
  select(Bricks)%>%
  gg_season( polar = FALSE)
```

```
## Plot variable not specified, automatically selected `y = Bricks`
```

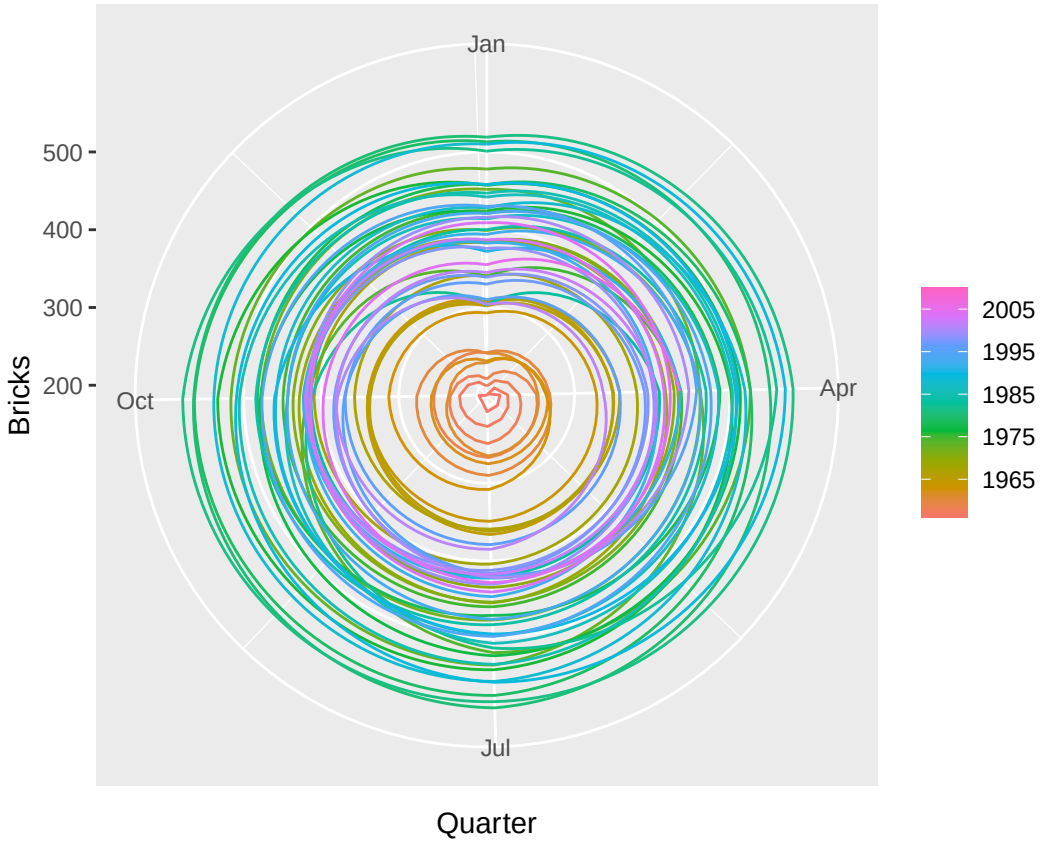
```
## Warning: Removed 20 rows containing missing values (`geom_line()`).
```




```
aus_production%>%
  select(Bricks)%>%
  gg_season( polar = TRUE)
```

```
## Plot variable not specified, automatically selected `y = Bricks`
```

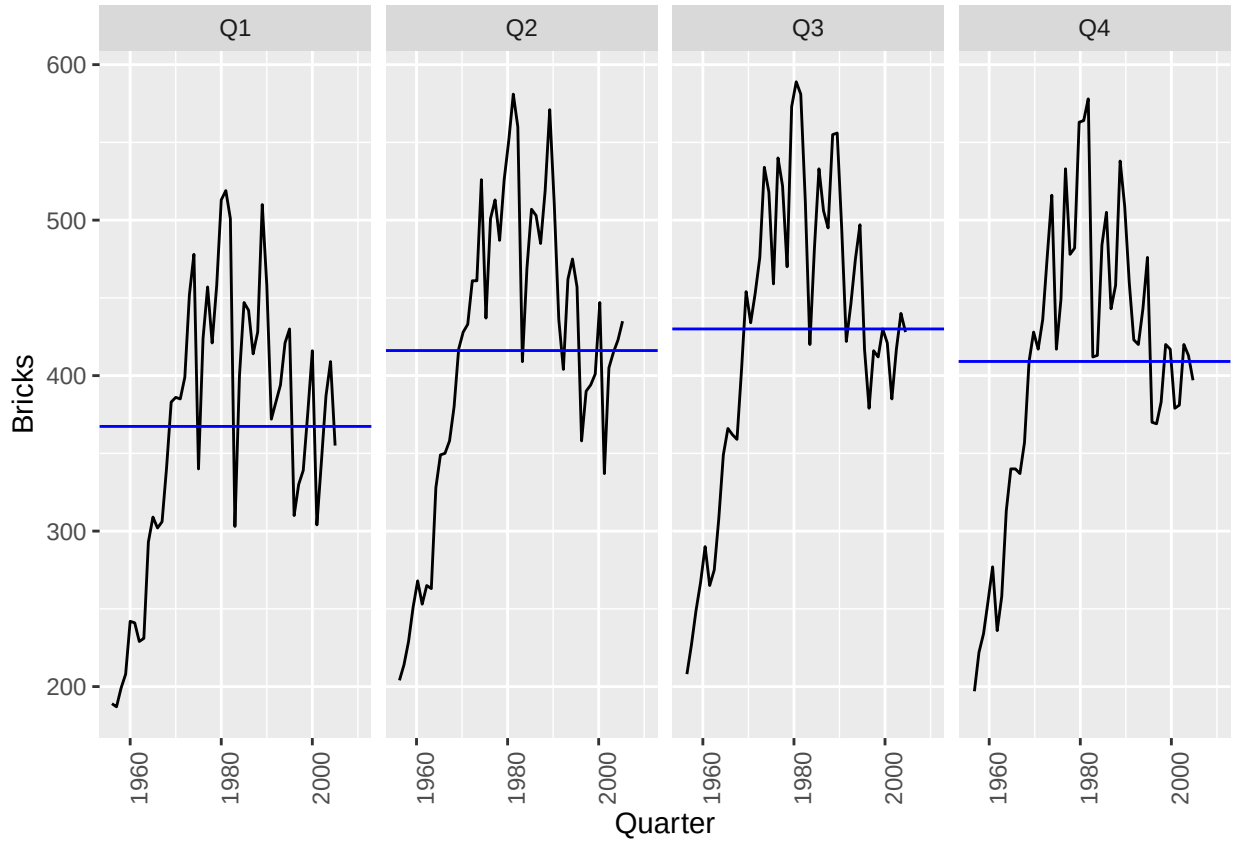
```
## Warning: Removed 20 rows containing missing values (`geom_line()`).
```




```
aus_production%>%
  select(Bricks)%>%
  gg_subseries()
```

```
## Plot variable not specified, automatically selected `y = Bricks`
```

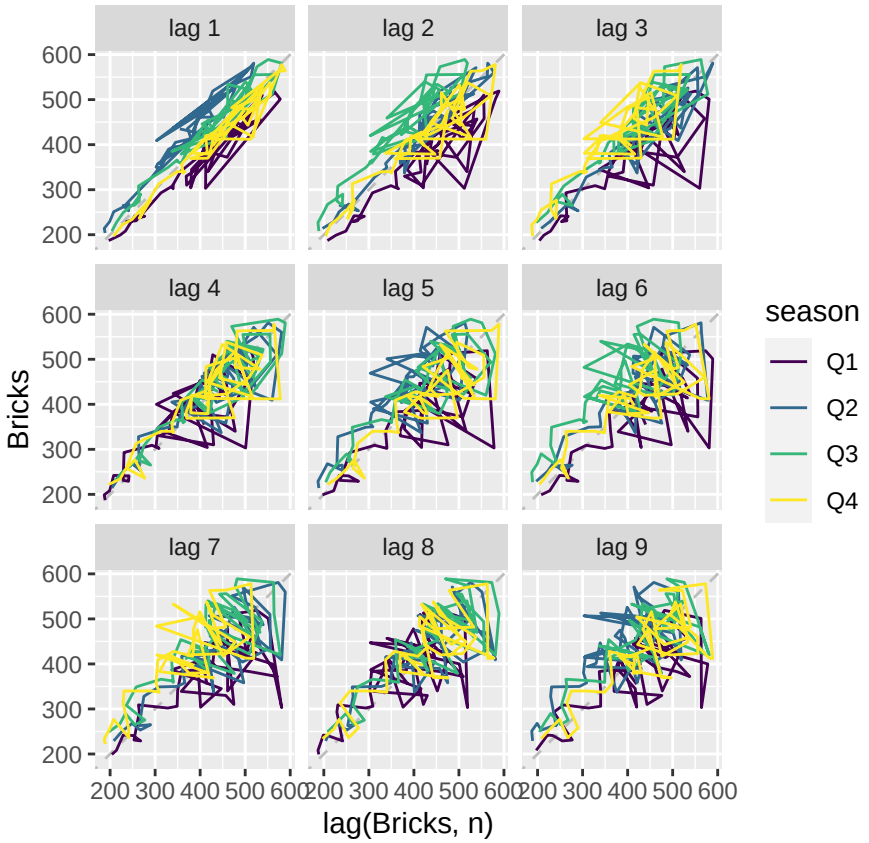
```
## Warning: Removed 5 rows containing missing values (`geom_line()`).
```



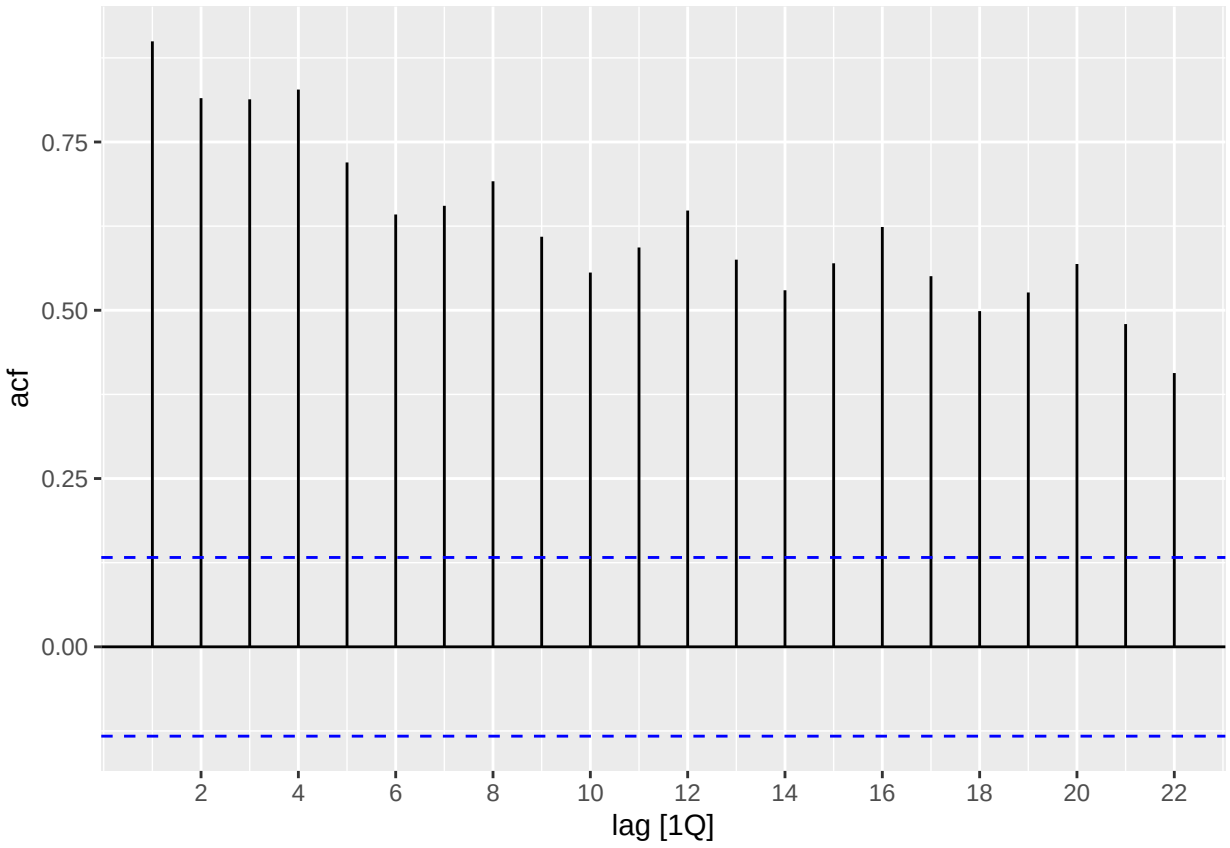

```
aus_production%>%
  select(Bricks)%>%
  gg_lag()
```

```
## Plot variable not specified, automatically selected `y = Bricks`
```

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




```
aus_production%>%
  select(Bricks)%>%
  ACF(aus_production$Bricks)%>%
  autoplot()
```



i

There is lots of cyclicity with frequent spikes and dips, but it does not appear to be consistent to a time period.There is a positive upward trend in the long term.

ii

The data being broken down to Quarters my influence how well we can assess the potential seasonality. As is, there does seem to be one.

iii

There seems to be some seasonality as far as Q1 and Q3 is concerned.

iv

The early 1980s has a significant dip so I would be curious to understand what may have cause this.

Hare

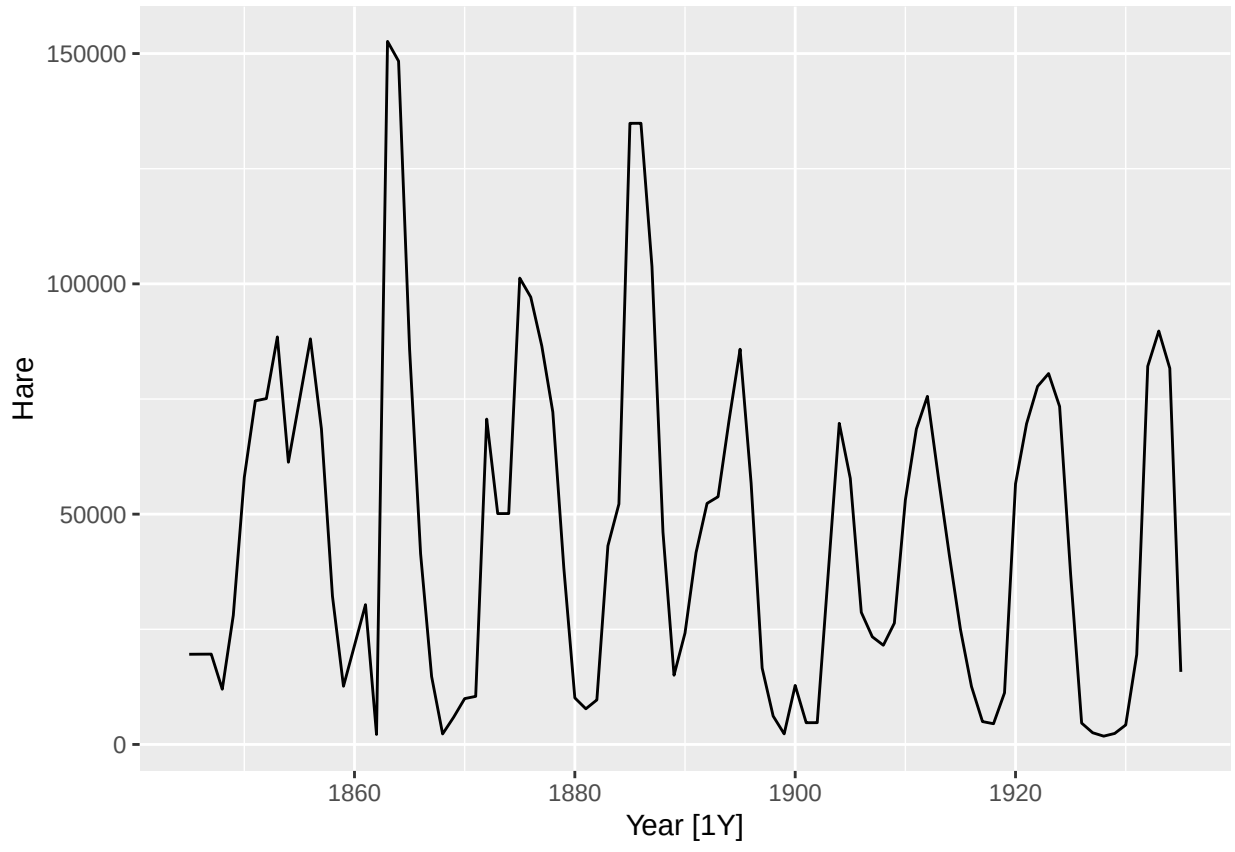
```
pelt
```

```
## # 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%>%
  select(Hare)%>%
  autoplot(period="year")
```

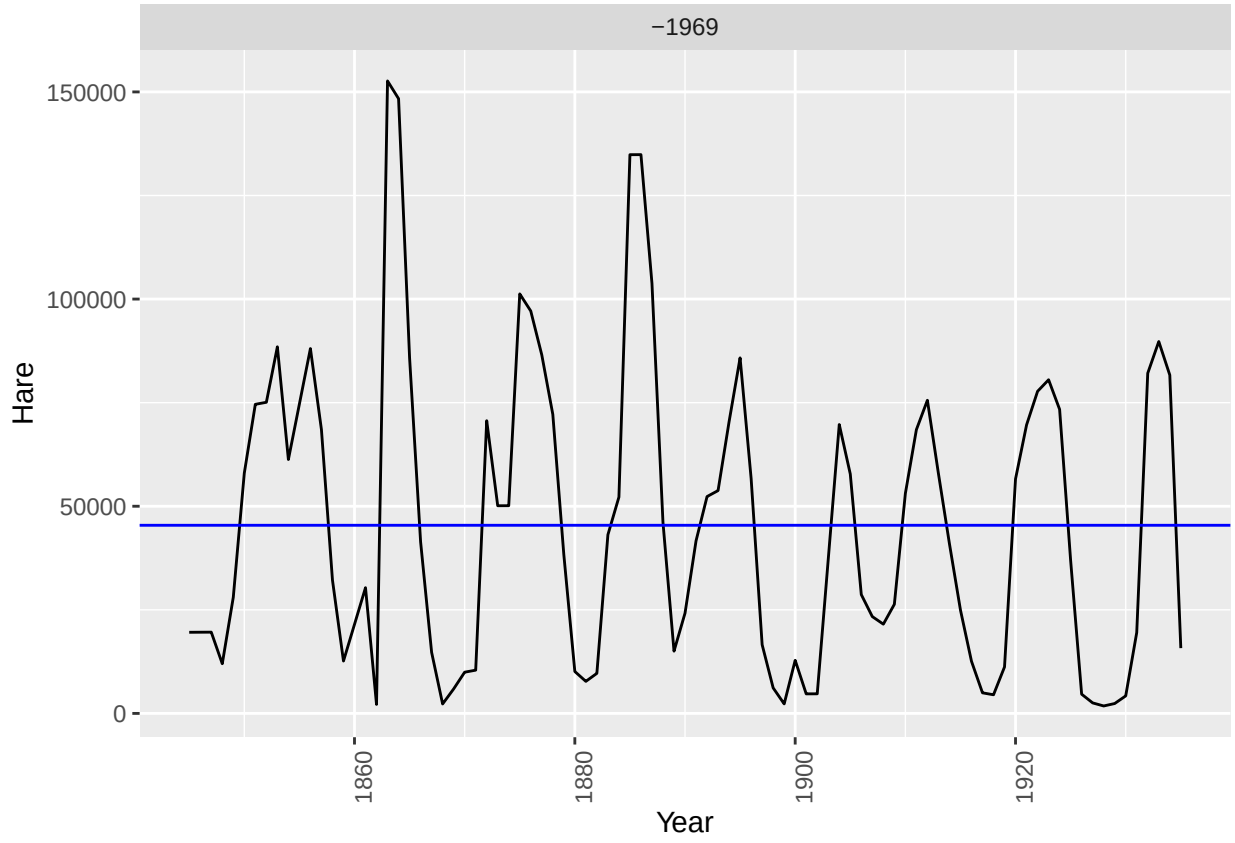
```
## Plot variable not specified, automatically selected `.vars = Hare`
```

```
## Warning in geom_line(...): Ignoring unknown parameters: `period`
```



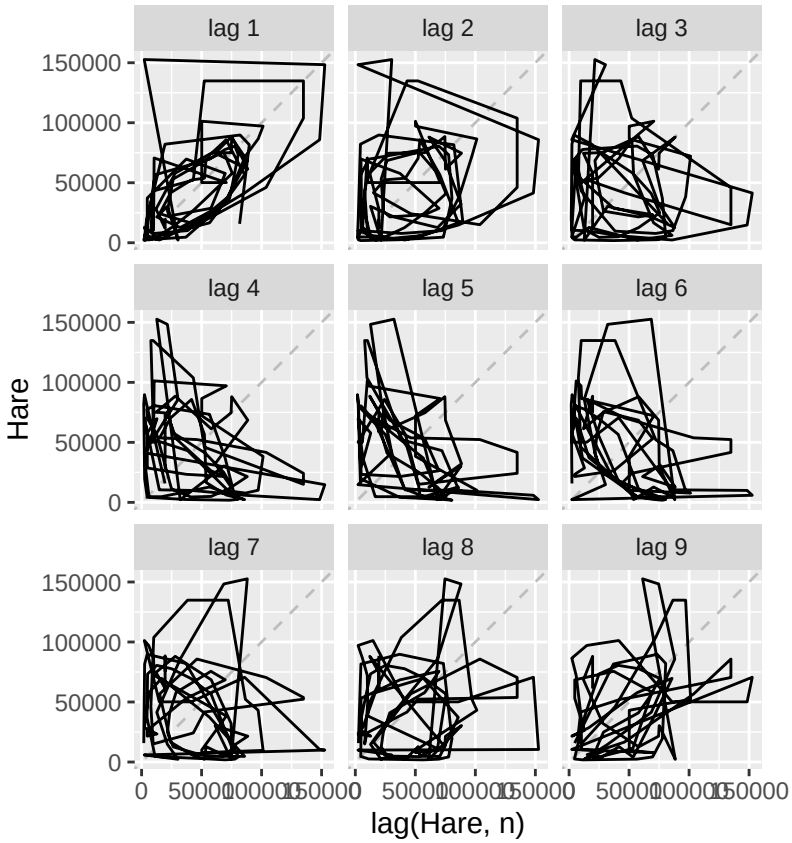
```
#Not possible
# pelt%>%
#   select(Hare)%>%
#   gg_season( polar = FALSE)
#
# pelt%>%
#   select(Hare)%>%
#   gg_season( polar = TRUE)

pelt%>%
  select(Hare)%>%
  gg_subseries()
```

```
## Plot variable not specified, automatically selected `y = Hare`
```



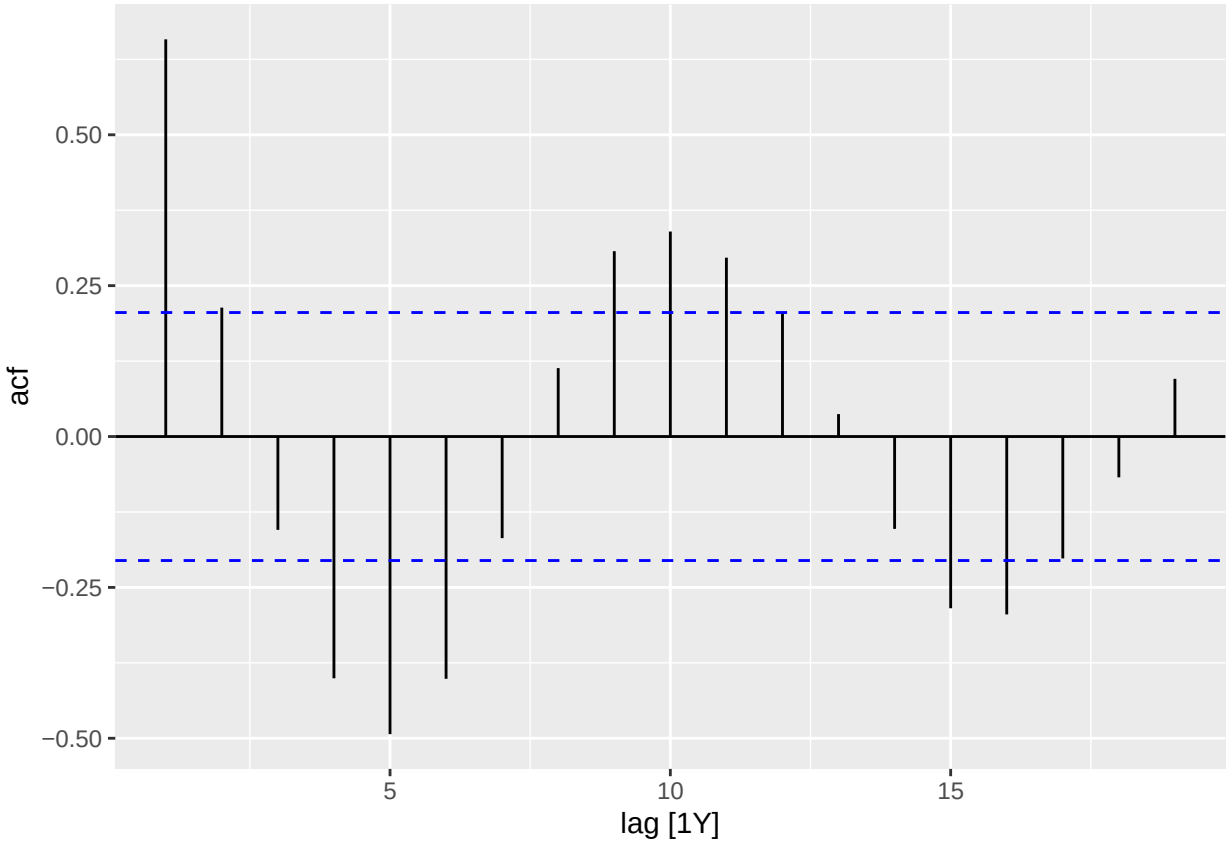

```
pelt%>%
  select(Hare)%>%
  gg_lag()
```

```
## Plot variable not specified, automatically selected `y = Hare`
```




```
pelt%>%
  select(Hare)%>%
  ACF()%>%
  autoplot()
```

```
## Response variable not specified, automatically selected `var = Hare`
```



i

The data is definitely cyclical but its not possible to teal seasonality since its at a annual basis.

ii

The data does not trend and varies a great deal. But seems to have a pattern at a 5 year interval.

iii

Again no seasonality

iv

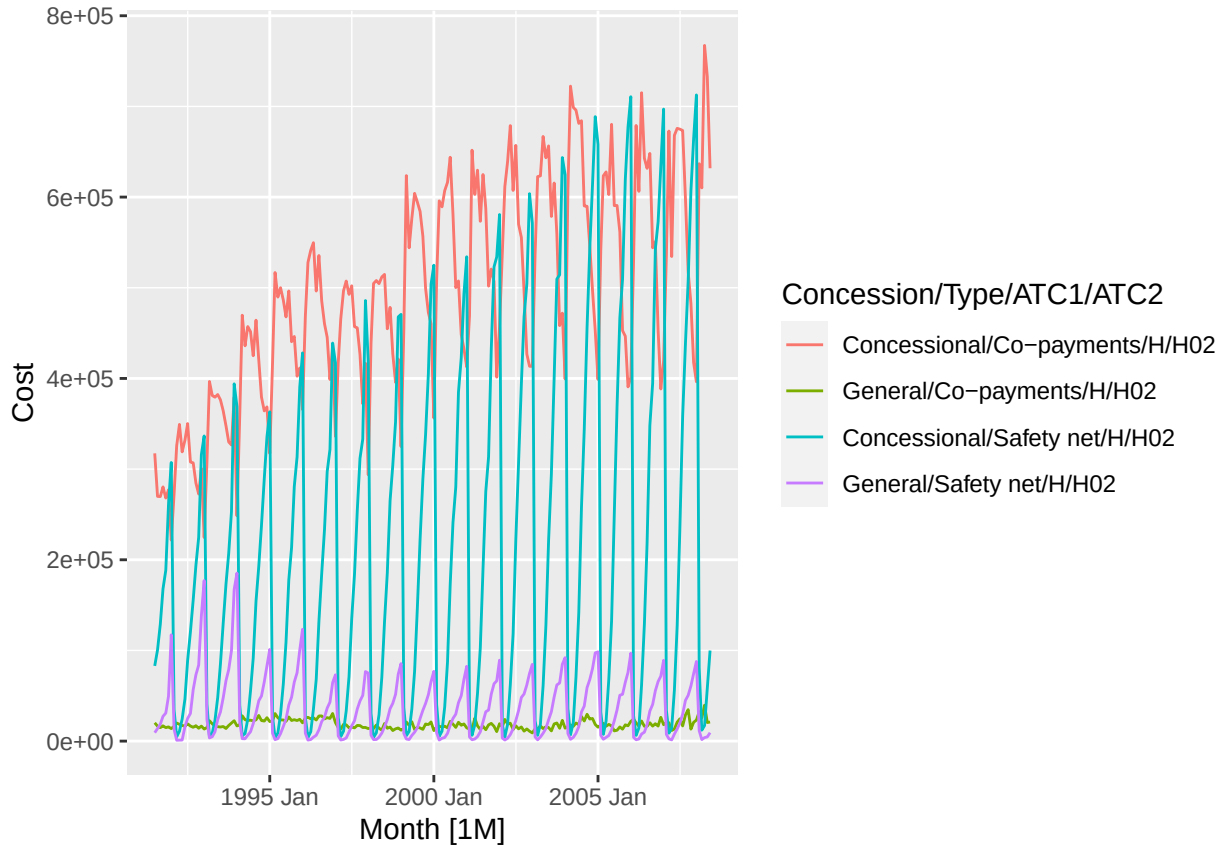
Im curious what caused the peak in the early 1860s

Cost

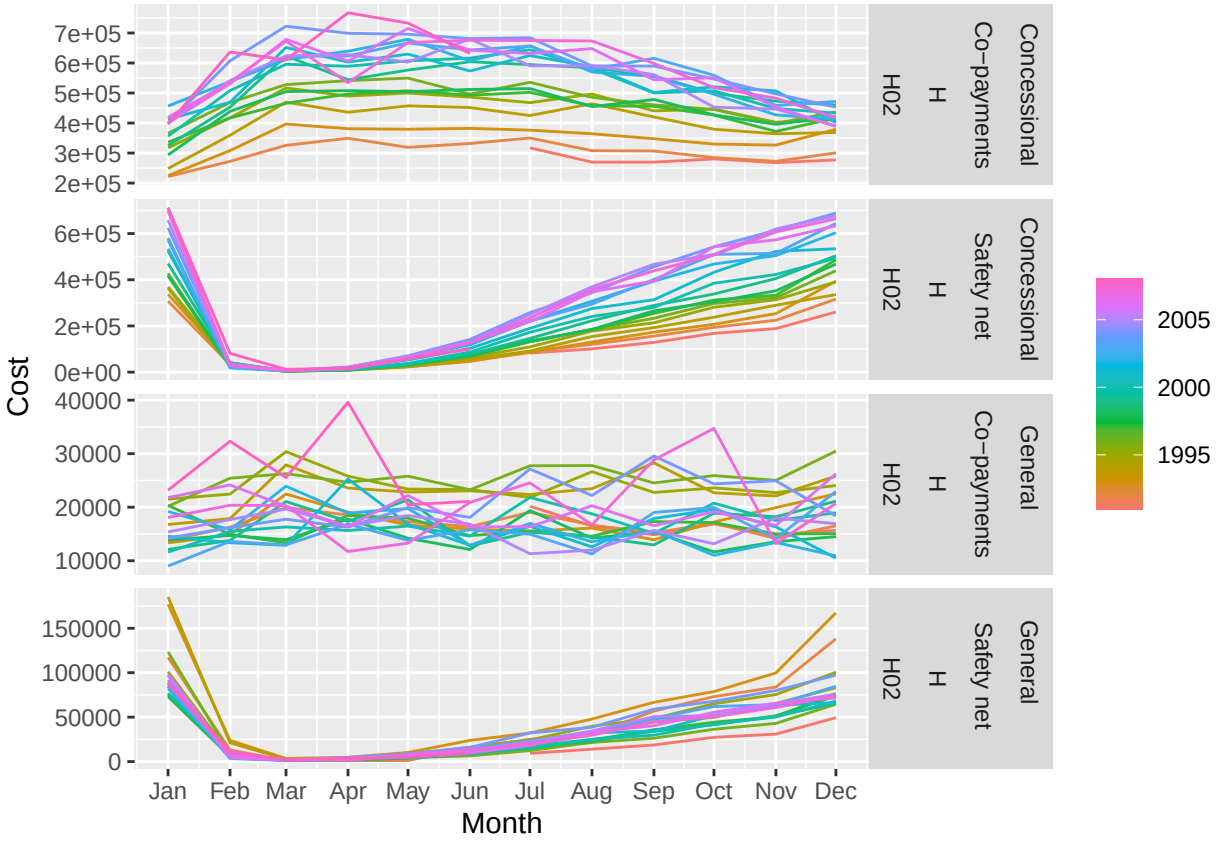
```
PBS
```

```
## # 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
```

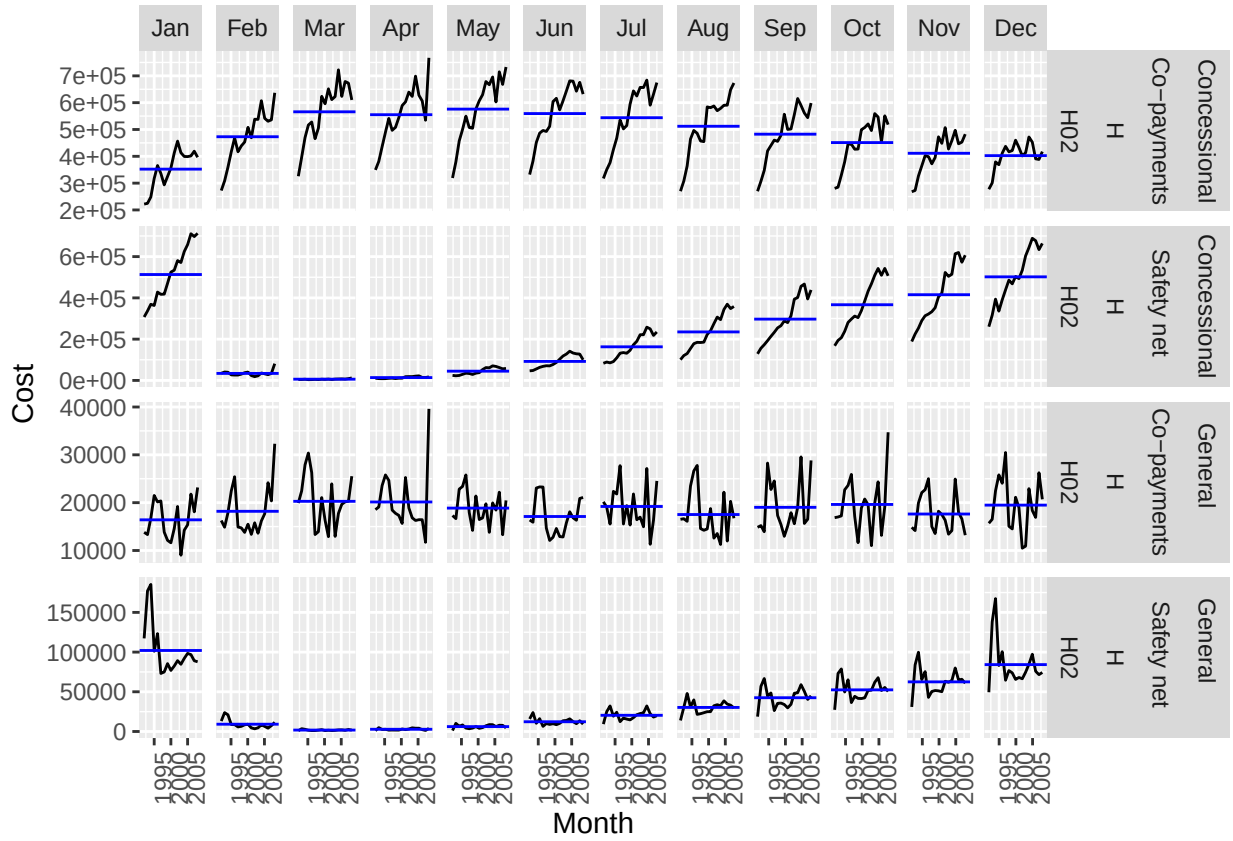
```
PBS%>%
  filter(ATC2=="H02")%>%
  autoplot(Cost)
```




```
PBS %>%
  filter(ATC2 == "H02") %>%
  gg_season(Cost, polar = FALSE)
```

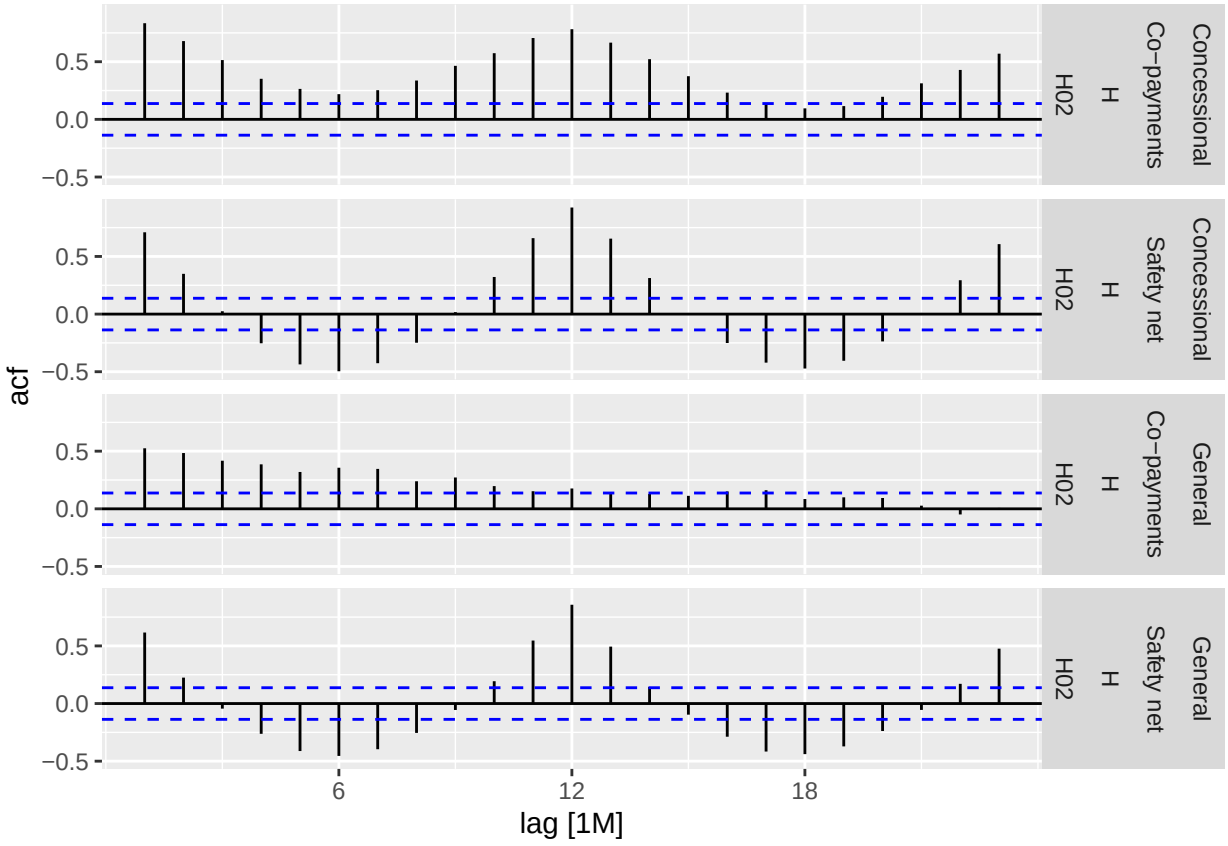



```
PBS %>%
  filter(ATC2 == "H02") %>%
  gg_subseries(Cost)
```




```
# PBS %>%
#   filter(ATC2 == "H02") %>%
#   gg_lag(Cost)
```

```
PBS %>%
  filter(ATC2 == "H02") %>%
  ACF(Cost)%>%
  autoplot()
```



i

The data is hard to interpret but it appears to trend upwards with cyclicity and seasonality. ### ii the data is very volatile but spikes mainly end of year it appears.

iii

The seasonality is at the end of the year.

iv

No year stands out, outside of the latest year having the highest cost.

Barrels

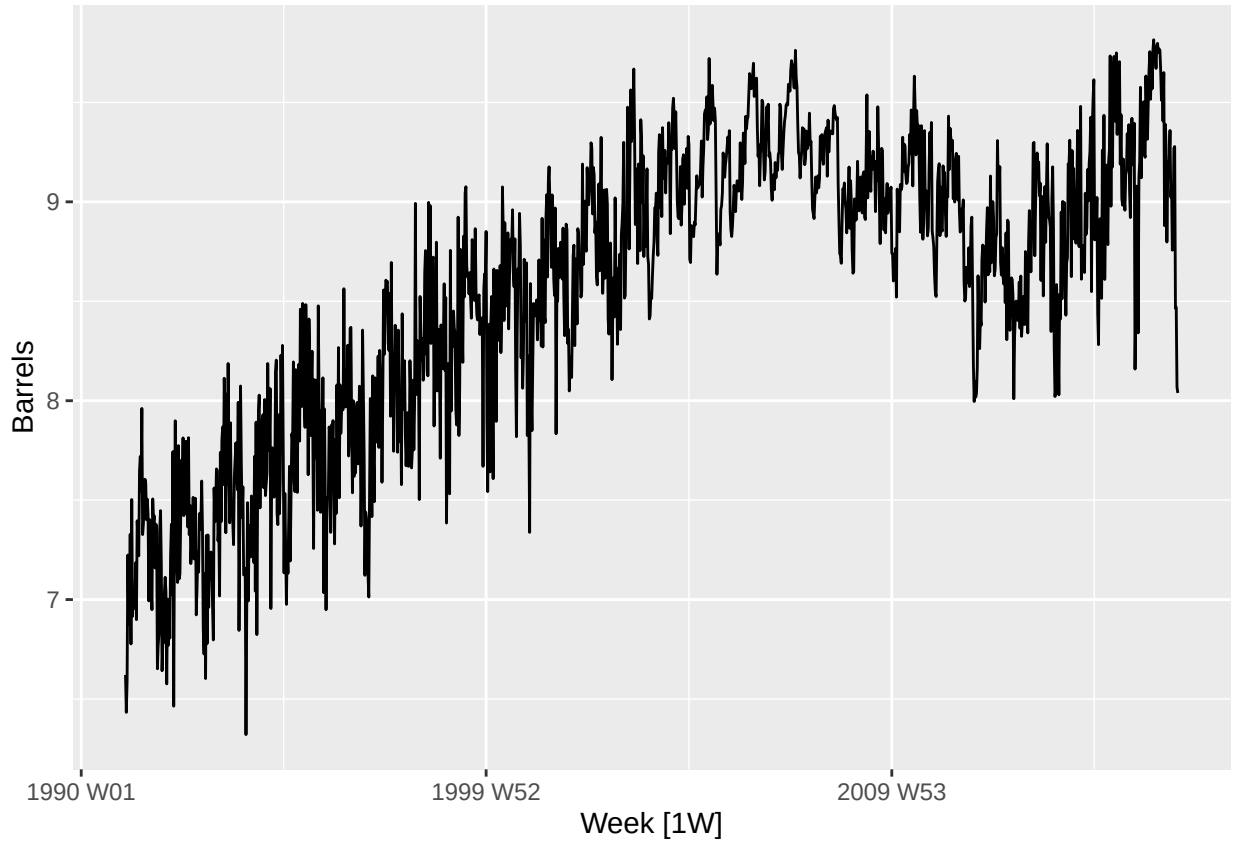
```
us_gasoline
```

```
## # 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
```

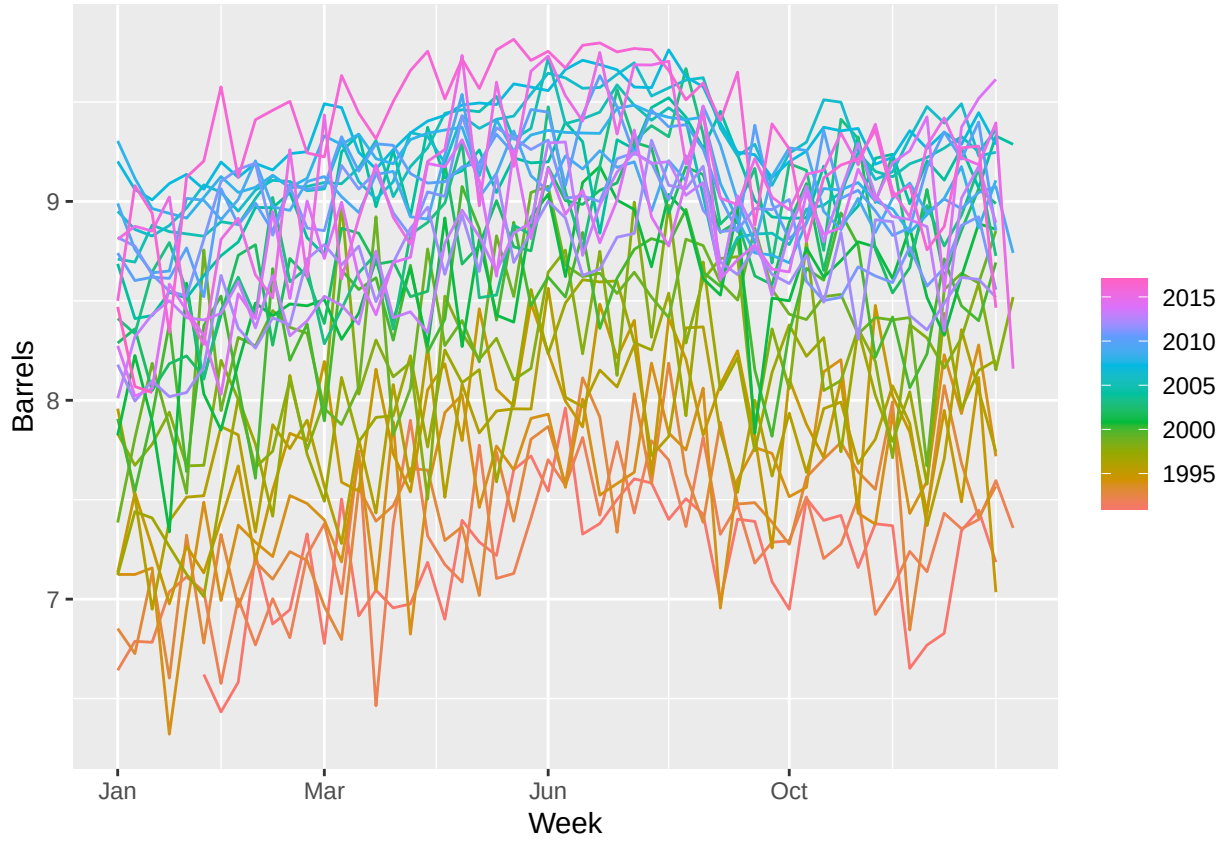
```
us_gasoline%>%
  select(Barrels)%>%
  autoplot()
```

```
## Plot variable not specified, automatically selected `.vars = Barrels`
```



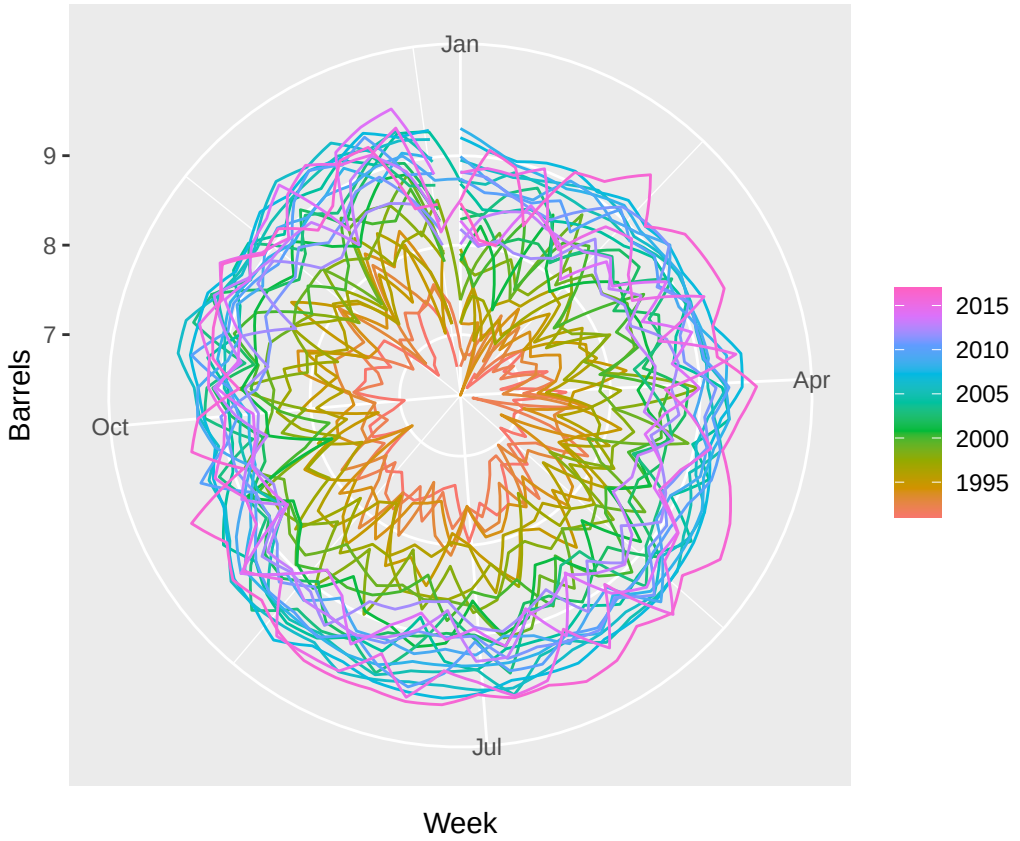

```
us_gasoline%>%
  select(Barrels)%>%
  gg_season(polar = FALSE)
```

```
## Plot variable not specified, automatically selected `y = Barrels`
```



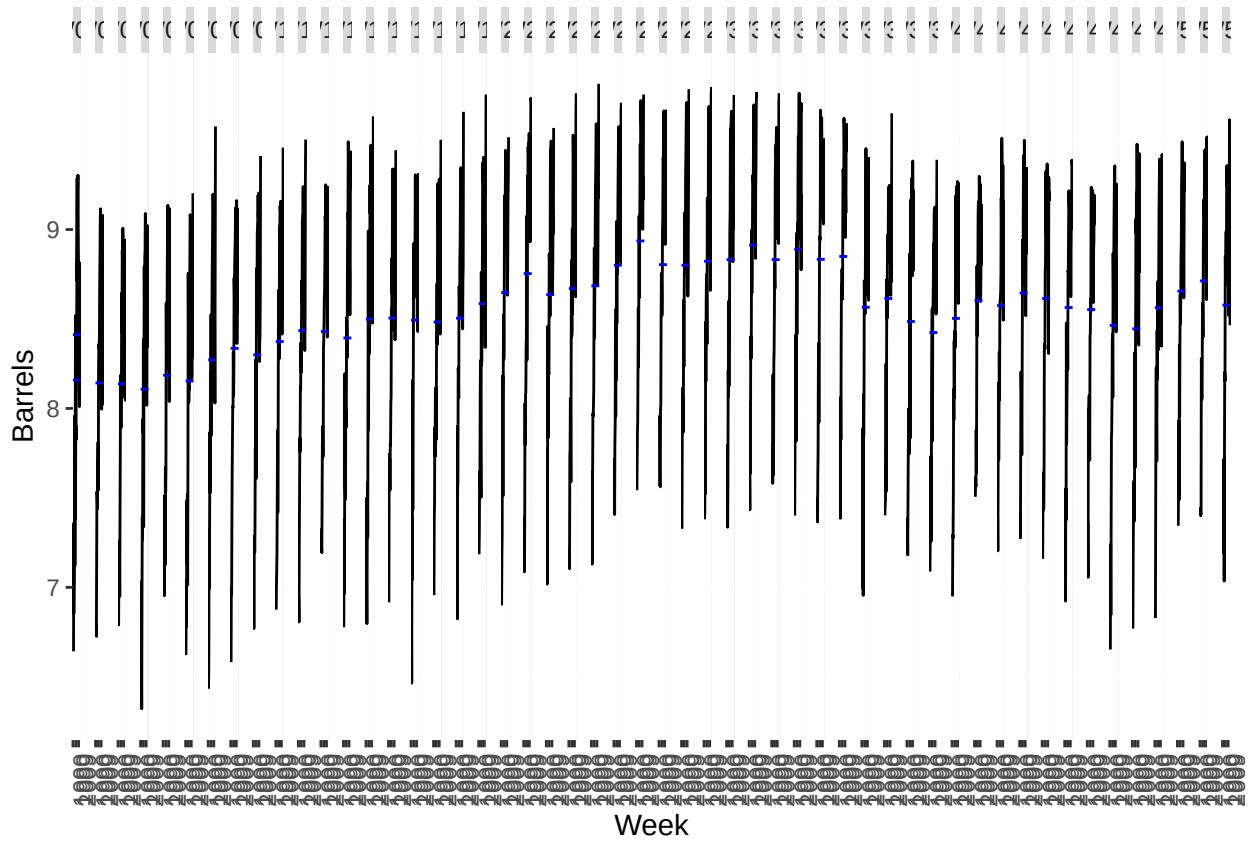

```
us_gasoline%>%
  select(Barrels)%>%
  gg_season( polar = TRUE)
```

```
## Plot variable not specified, automatically selected `y = Barrels`
```



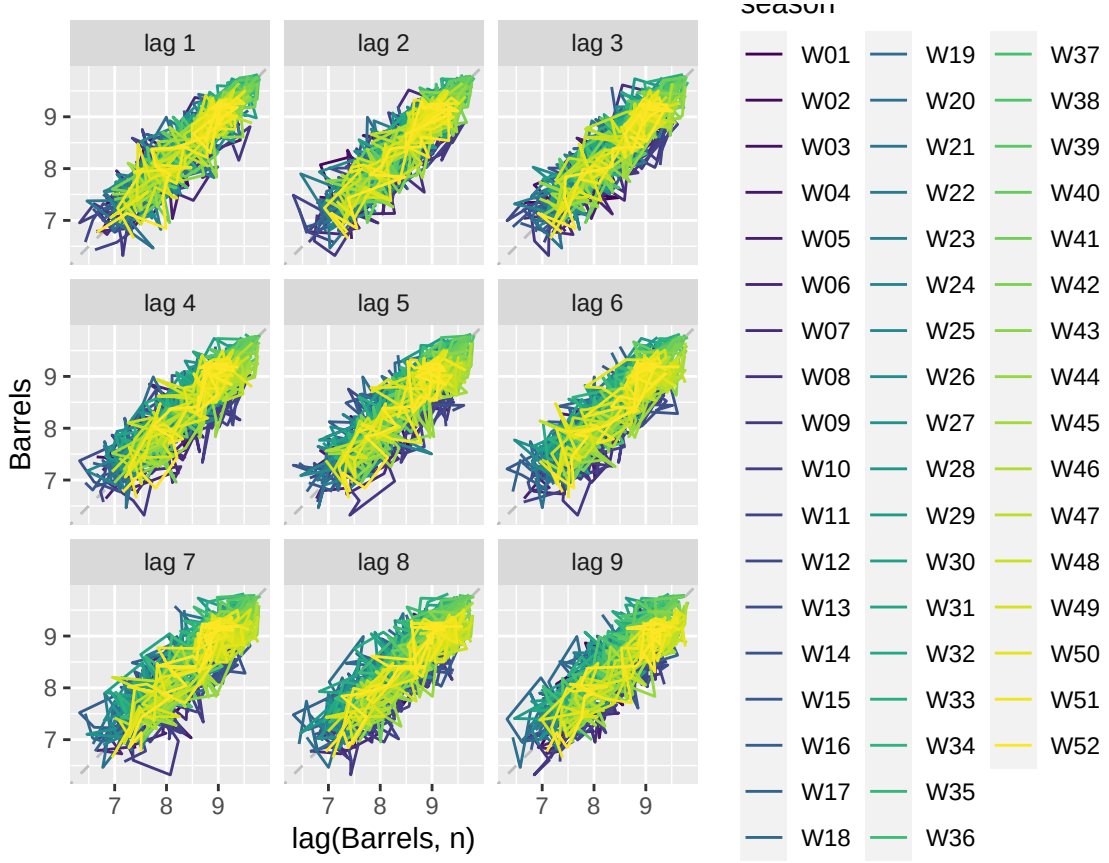

```
us_gasoline%>%
  select(Barrels)%>%
  gg_subseries()
```

```
## Plot variable not specified, automatically selected `y = Barrels`
```



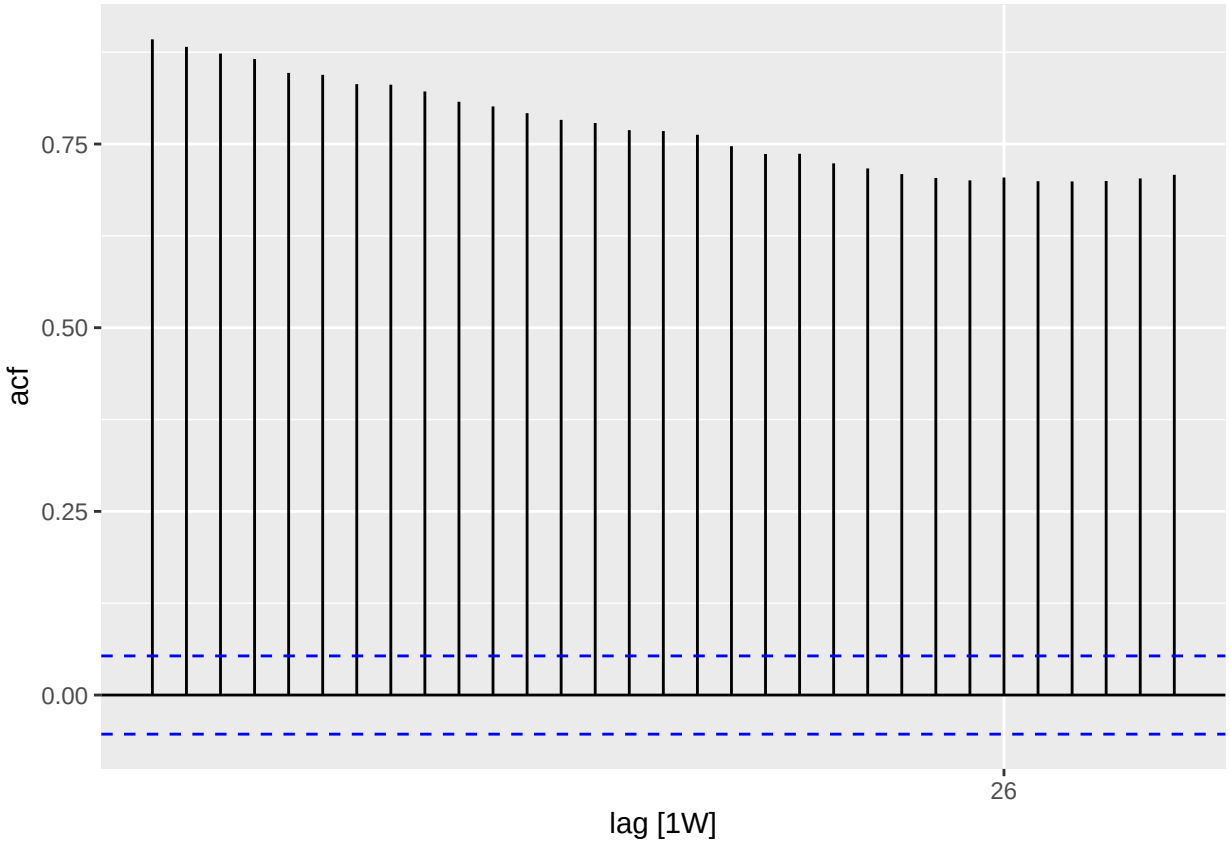

```
us_gasoline%>%
  select(Barrels)%>%
  gg_lag()
```

```
## Plot variable not specified, automatically selected `y = Barrels`
```




```
us_gasoline %>%
  select(Barrels) %>%
  ACF()%>%
  autoplot()
```

```
## Response variable not specified, automatically selected `var = Barrels`
```



i

Primarily an upward trend with a dip near the most recent year

ii

Its possible the barrels value is impacted by supply.

iii

There does not appear to be seasonality of cyclicity

iv

The most recent dip is interesting and I wonder if its just a data collection issue.