Data 624_Exercise 3.7_HW2

Jiaxin Zheng

2025-02-17

3.7 Exercises:

```r
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()
```

```r
library(seasonal)
```

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

```
##
## Attaching package: 'seasonal'
```

```
## The following object is masked from 'package:tibble':
##
##     view
```

1.Consider the GDP information in global_economy. Plot the GDP per capita for each country over time. Which country has the highest GDP per capita? How has this changed over time?

- Monaco is the highest GDP per capita country in 2014. Overall the country's GDP per Capita is increasing, except around the ear of 2000, it dropped.

```r
?global_economy
```

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

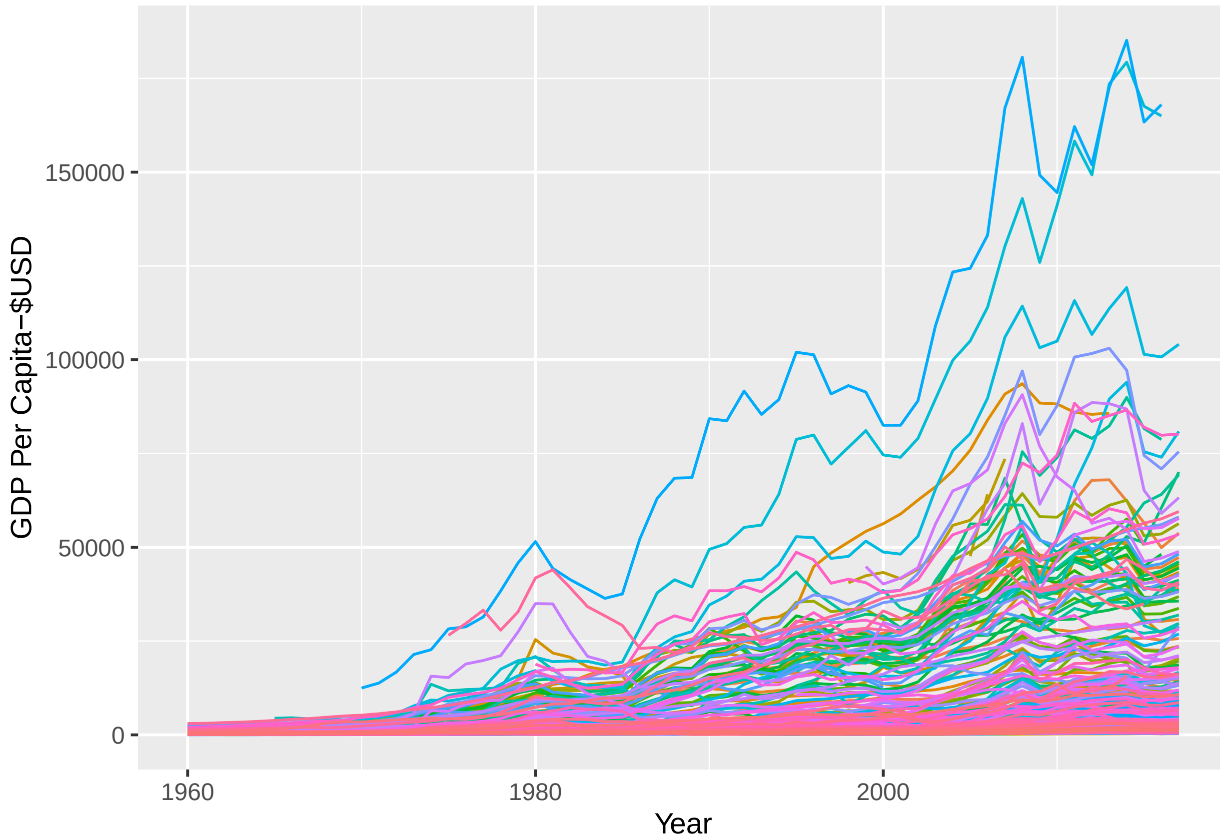
```r
head(global_economy)
```

```
## # A tsibble: 6 x 9 [1Y]
## # Key:       Country [1]
##   Country     Code   Year         GDP Growth   CPI Imports Exports Population
##   <fct>       <fct> <dbl>       <dbl>  <dbl> <dbl>   <dbl>   <dbl>      <dbl>
## 1 Afghanistan AFG    1960  537777811.     NA    NA    7.02    4.13    8996351
## 2 Afghanistan AFG    1961  548888896.     NA    NA    8.10    4.45    9166764
## 3 Afghanistan AFG    1962  546666678.     NA    NA    9.35    4.88    9345868
## 4 Afghanistan AFG    1963  751111191.     NA    NA   16.9     9.17    9533954
## 5 Afghanistan AFG    1964  800000044.     NA    NA   18.1     8.89    9731361
## 6 Afghanistan AFG    1965 1006666638.     NA    NA   21.4    11.3     9938414
```

```r
global_economy <- global_economy %>%
  mutate(GDP_cap = GDP/Population)

global_economy %>%
  autoplot(GDP_cap,show.legend = FALSE) +
  labs(x = "Year", y = "GDP Per Capita-$USD")
```

```
## Warning: Removed 3242 rows containing missing values or values outside the scale range
## (`geom_line()`).
```

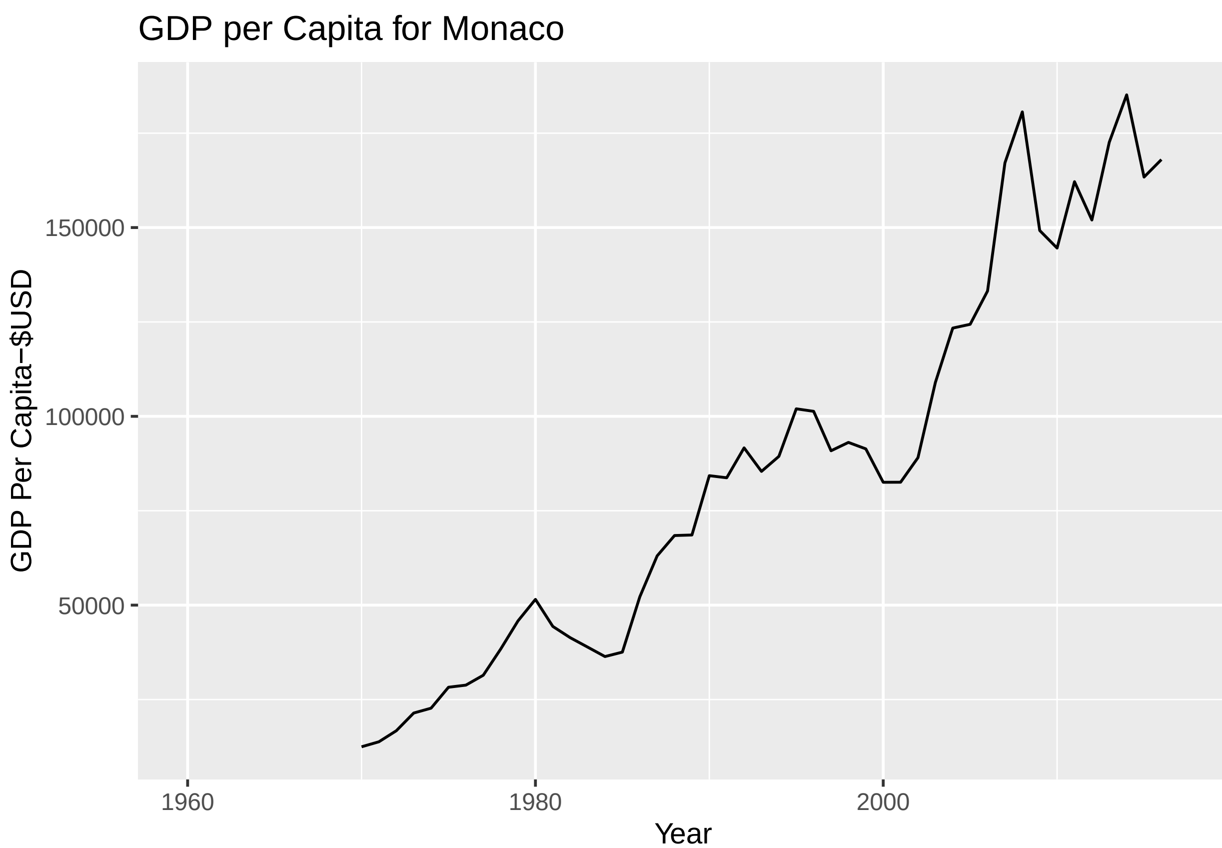



```r
global_economy %>%
  filter(GDP_cap == max(GDP_cap, na.rm=T))
```

```
## # A tsibble: 1 x 10 [1Y]
## # Key:       Country [1]
##   Country Code   Year        GDP Growth   CPI Imports Exports Population GDP_cap
##   <fct>   <fct> <dbl>      <dbl>  <dbl> <dbl>   <dbl>   <dbl>      <dbl>   <dbl>
## 1 Monaco  MCO    2014     7.06e9   7.18    NA      NA      NA      38132 185153.
```

```r
global_economy %>%
  filter(Country == "Monaco") %>%
  autoplot(GDP_cap)+
  labs(x = "Year", y = "GDP Per Capita-$USD", title = "GDP per Capita for Monaco")
```

```
## Warning: Removed 11 rows containing missing values or values outside the scale range
## (‘geom_line()‘).
```



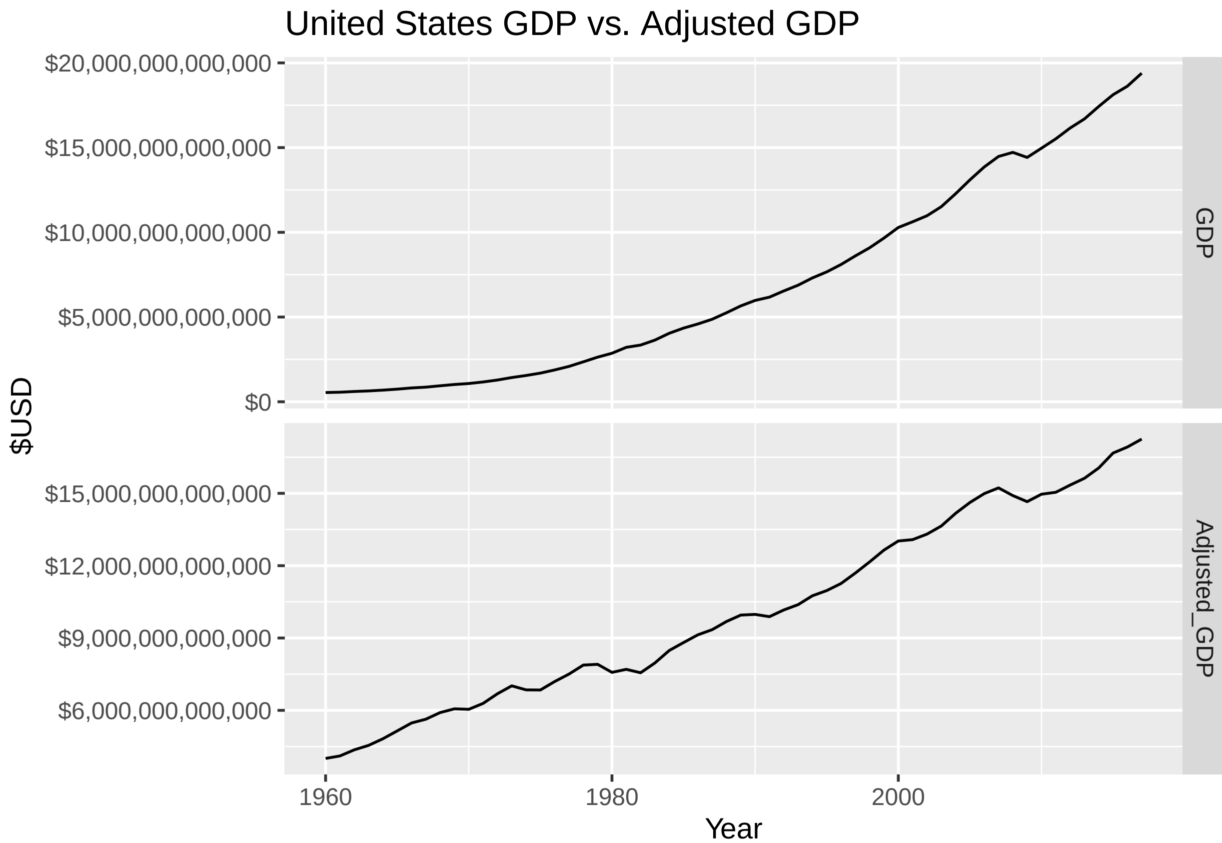
2.For each of the following series, make a graph of the data. If transforming seems appropriate, do so and describe the effect.

- a.United States GDP from global_economy.
- a Answer: There is not much change between the United States GDP vs. Adjusted GDP. The trends are going upward.

```
us_economy <- global_economy %>%
  filter(Country == "United States") %>%
  index_by(Year)

us_economy <- us_economy %>%
  mutate(Adjusted_GDP = GDP / CPI * 100) %>%
  pivot_longer(c(GDP, Adjusted_GDP), names_to = "GDP_Type", values_to = "GDP_Value") %>%
  mutate(GDP_Type = factor(GDP_Type, levels = c("GDP", "Adjusted_GDP")))

ggplot(us_economy, aes(x = Year, y = GDP_Value)) +
  geom_line() +
  facet_grid(GDP_Type ~ ., scales = "free_y") +
  labs(title = "United States GDP vs. Adjusted GDP",
       y = "$USD",
       x = "Year") +
  scale_y_continuous(labels = scales::dollar)
```



- b.Slaughter of Victorian "Bulls, bullocks and steers" in aus_livestock.
- b Answer: The data for "Bulls, bullocks and steers" in aus_livestock for State Victoria doesn't appear any upward trend, but it shows seasonal cyclicity. The lambda guerrero is a negative number, I don't think the transformation is useful.

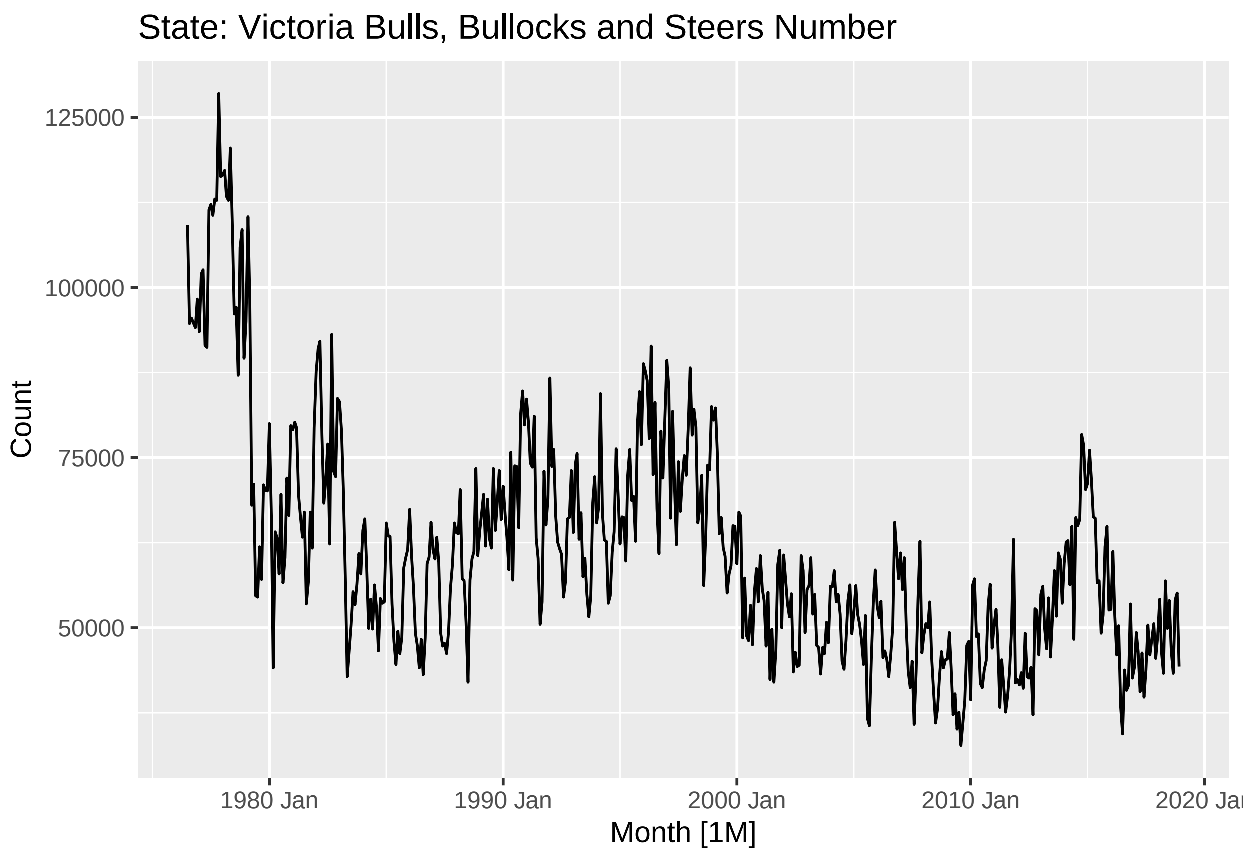
```
?aus_livestock
```

```
head(aus_livestock)
```

```
## # A tsibble: 6 x 4 [1M]
## # Key:       Animal, State [1]
##      Month Animal                     State                        Count
##      <mth> <fct>                      <fct>                        <dbl>
## 1 1976 Jul Bulls, bullocks and steers Australian Capital Territory  2300
## 2 1976 Aug Bulls, bullocks and steers Australian Capital Territory  2100
## 3 1976 Sep Bulls, bullocks and steers Australian Capital Territory  2100
## 4 1976 Oct Bulls, bullocks and steers Australian Capital Territory  1900
## 5 1976 Nov Bulls, bullocks and steers Australian Capital Territory  2100
## 6 1976 Dec Bulls, bullocks and steers Australian Capital Territory  1800
```

```
aus_livestock %>%
  filter(Animal == "Bulls, bullocks and steers", State == "Victoria") %>%
  autoplot() +
  labs(title = "State: Victoria Bulls, Bullocks and Steers Number")
```

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




```r
lambda <- aus_livestock %>%
  filter(Animal == "Bulls, bullocks and steers", State == "Victoria") %>%
  features(Count,features = guerrero)

lambda
```

```
## # A tibble: 1 x 3
##   Animal                     State    lambda_guerrero
##   <fct>                      <fct>              <dbl>
## 1 Bulls, bullocks and steers Victoria         -0.0446
```

- c.Victorian Electricity Demand from vic_elec.
- c Answer: I don't think the transformation is useful. Seems there is no upward or downward trend, but a seasonal pattern.

```r
?vic_elec
```

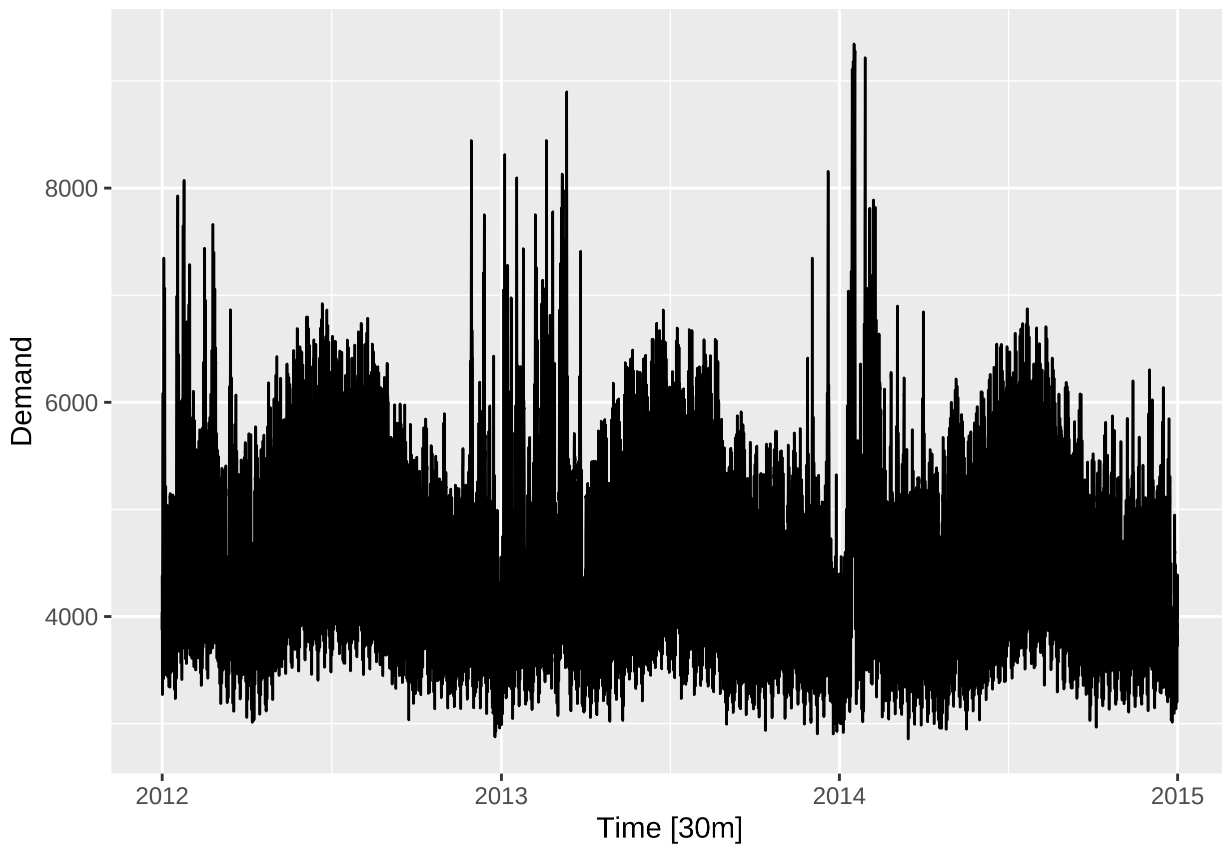
```r
head(vic_elec)
```

```
## # A tsibble: 6 x 5 [30m] <Australia/Melbourne>
##   Time                Demand Temperature Date       Holiday
##   <dttm>               <dbl>       <dbl> <date>     <lgl>
```

```
## 1 2012-01-01 00:00:00  4383.        21.4 2012-01-01 TRUE
## 2 2012-01-01 00:30:00  4263.        21.0 2012-01-01 TRUE
## 3 2012-01-01 01:00:00  4049.        20.7 2012-01-01 TRUE
## 4 2012-01-01 01:30:00  3878.        20.6 2012-01-01 TRUE
## 5 2012-01-01 02:00:00  4036.        20.4 2012-01-01 TRUE
## 6 2012-01-01 02:30:00  3866.        20.2 2012-01-01 TRUE
```

```
vic_elec %>%
  autoplot()
```

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



- d.Gas production from aus_production.
- d Answer: The data shows an upward trend so the transformation is useful. Compared to the original data plot, a Box-Cox transformation helps us to see the seasonal trend better.

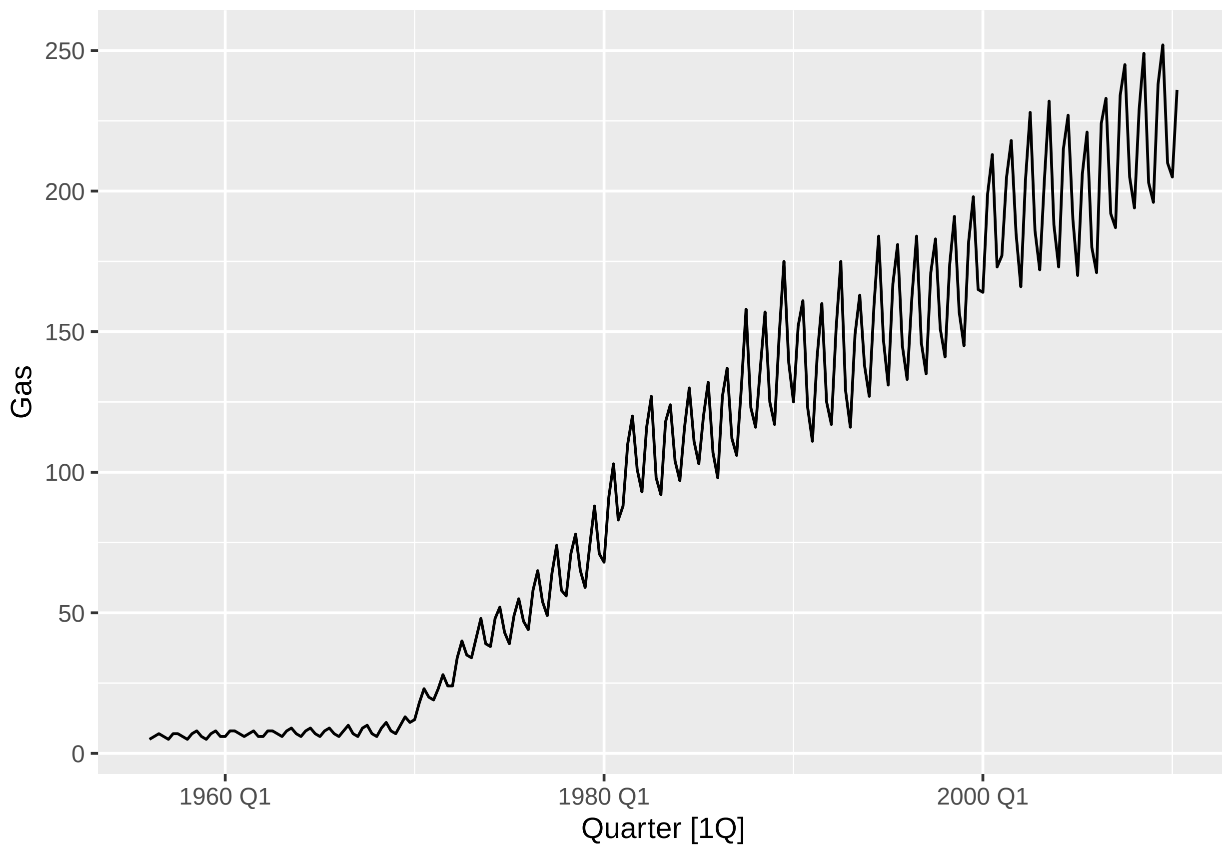
```
?aus_production
```

```
head(aus_production)
```

```
## # A tsibble: 6 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
```

```
aus_production %>%
  autoplot(Gas)
```



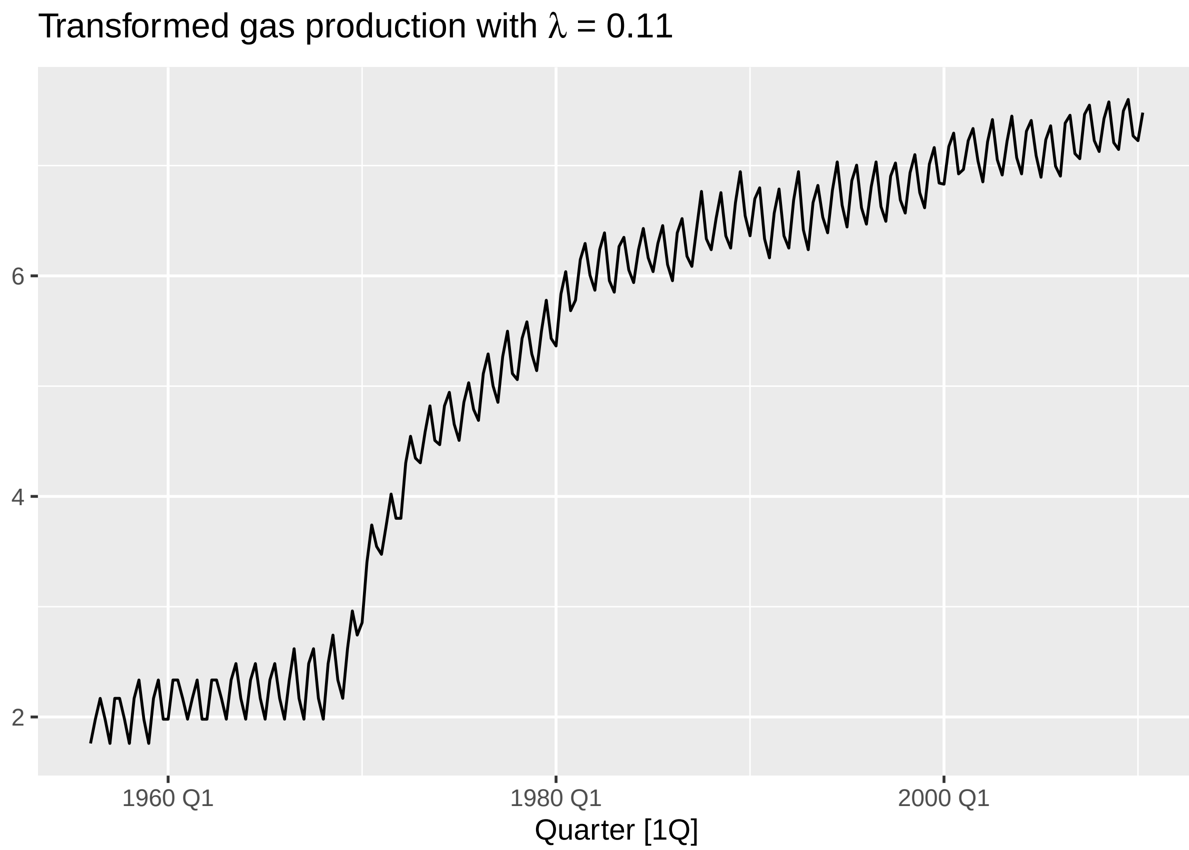

```
lambda_2 <- aus_production %>%
  features(Gas,features = guerrero)

lambda_2
```

```
## # A tibble: 1 x 1
##   lambda_guerrero
##             <dbl>
## 1           0.110
```

```
aus_production %>%
  autoplot(box_cox(Gas, lambda_2)) +
  labs(y = "",
       title = latex2exp::TeX(paste0(
```

```r
       "Transformed gas production with $\\lambda$ = ",
       round(lambda_2,2))))
```



3. Why is a Box-Cox transformation unhelpful for the canadian_gas data?

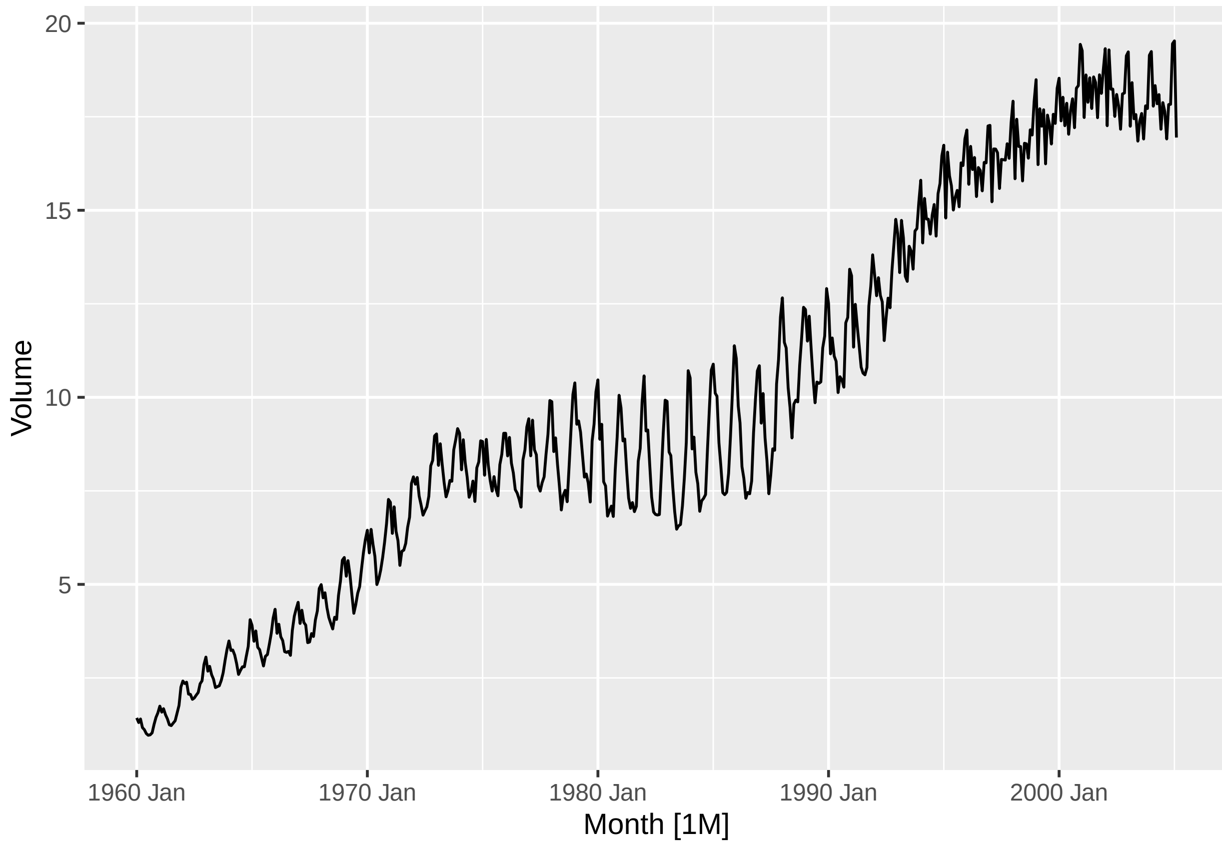
- Answer: It could be the data's trend is already spread well.

```r
?canadian_gas
```

```r
head(canadian_gas)
```

```
## # A tsibble: 6 x 2 [1M]
##       Month Volume
##       <mth>  <dbl>
## 1 1960 Jan   1.43
## 2 1960 Feb   1.31
## 3 1960 Mar   1.40
## 4 1960 Apr   1.17
## 5 1960 May   1.12
## 6 1960 Jun   1.01
```

```r
canadian_gas %>%
  autoplot(Volume)
```



4.What Box-Cox transformation would you select for your retail data (from Exercise 7 in Section 2.10)?

- Answer: The lambda is 0.08303631, the BOX-Cox transformation may be the best for this retail data?

```r
head(aus_retail)
```

```
## # A tsibble: 6 x 5 [1M]
## # Key:       State, Industry [1]
##   State                        Industry            `Series ID`    Month Turnover
##   <chr>                        <chr>               <chr>          <mth>    <dbl>
## 1 Australian Capital Territory Cafes, restaurants~ A3349849A   1982 Apr      4.4
## 2 Australian Capital Territory Cafes, restaurants~ A3349849A   1982 May      3.4
## 3 Australian Capital Territory Cafes, restaurants~ A3349849A   1982 Jun      3.6
## 4 Australian Capital Territory Cafes, restaurants~ A3349849A   1982 Jul      4
## 5 Australian Capital Territory Cafes, restaurants~ A3349849A   1982 Aug      3.6
## 6 Australian Capital Territory Cafes, restaurants~ A3349849A   1982 Sep      4.2
```

```r
set.seed(12345678)
myseries <- aus_retail %>%
```

```r
  filter(`Series ID` == sample(aus_retail$`Series ID`,1)) %>%
  features(Turnover, features = guerrero) %>%
  pull(lambda_guerrero)

myseries
```

```
## [1] 0.08303631
```

5.For the following series, find an appropriate Box-Cox transformation in order to stabilise the variance. Tobacco from aus_production, Economy class passengers between Melbourne and Sydney from ansett, and Pedestrian counts at Southern Cross Station from pedestrian.

- 5a Answer: Tobacco from aus_production: lambda = 0.9264636

```r
print(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
```

```r
aus_production %>%
  features(Tobacco,features = guerrero) %>%
  pull(lambda_guerrero)
```

```
## [1] 0.9264636
```

- 5b Answer: Economy class passengers between Melbourne and Sydney from ansett: lambda = 1.999927

```r
head(ansett)
```

```
## # A tsibble: 6 x 4 [1W]
## # Key:       Airports, Class [1]
##       Week Airports Class    Passengers
##     <week> <chr>    <chr>         <dbl>
## 1 1989 W28 ADL-PER  Business        193
## 2 1989 W29 ADL-PER  Business        254
## 3 1989 W30 ADL-PER  Business        185
## 4 1989 W31 ADL-PER  Business        254
## 5 1989 W32 ADL-PER  Business        191
## 6 1989 W33 ADL-PER  Business        136
```

```r
ansett %>%
  filter(Class == "Economy", Airports == "MEL-SYD") %>%
  features(Passengers,features = guerrero) %>%
  pull(lambda_guerrero)
```

```
## [1] 1.999927
```

- 5c: Pedestrian counts at Southern Cross Station from pedestrian: lambda = -0.2501616

```r
head(pedestrian)
```

```
## # A tsibble: 6 x 5 [1h] <Australia/Melbourne>
## # Key:       Sensor [1]
##   Sensor         Date_Time           Date        Time Count
##   <chr>          <dttm>              <date>     <int> <int>
## 1 Birrarung Marr 2015-01-01 00:00:00 2015-01-01     0  1630
## 2 Birrarung Marr 2015-01-01 01:00:00 2015-01-01     1   826
## 3 Birrarung Marr 2015-01-01 02:00:00 2015-01-01     2   567
## 4 Birrarung Marr 2015-01-01 03:00:00 2015-01-01     3   264
## 5 Birrarung Marr 2015-01-01 04:00:00 2015-01-01     4   139
## 6 Birrarung Marr 2015-01-01 05:00:00 2015-01-01     5    77
```

```r
pedestrian %>%
  filter(Sensor == "Southern Cross Station") %>%
  features(Count,features = guerrero) %>%
  pull(lambda_guerrero)
```

```
## [1] -0.2501616
```

7.Consider the last five years of the Gas data from aus_production.

- a.Plot the time series. Can you identify seasonal fluctuations and/or a trend-cycle?
- Answer: The overall trend is going upward, every 3rd quarter is the highest, and every 1st quarter is the lowest.

```r
?aus_production
```

```r
head(aus_production)
```

```
## # A tsibble: 6 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
```

```r
gas <- tail(aus_production, 5*4) |> select(Gas)
```

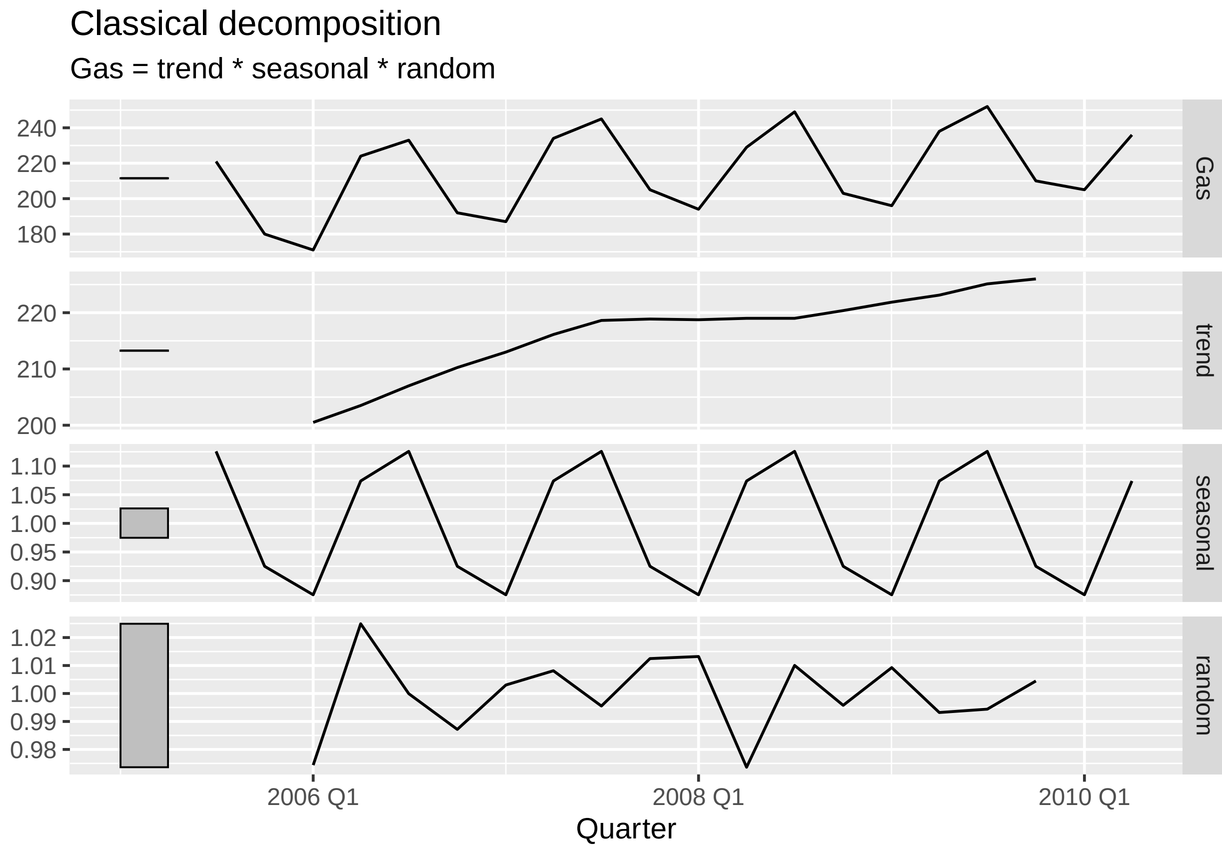
```r
gas %>%
  autoplot()
```

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

- b.Use classical_decomposition with type=multiplicative to calculate the trend-cycle and seasonal indices.

```r
gas %>%
  model(classical_decomposition(Gas, type = "multiplicative")) %>%
  components() %>%
  autoplot()
```

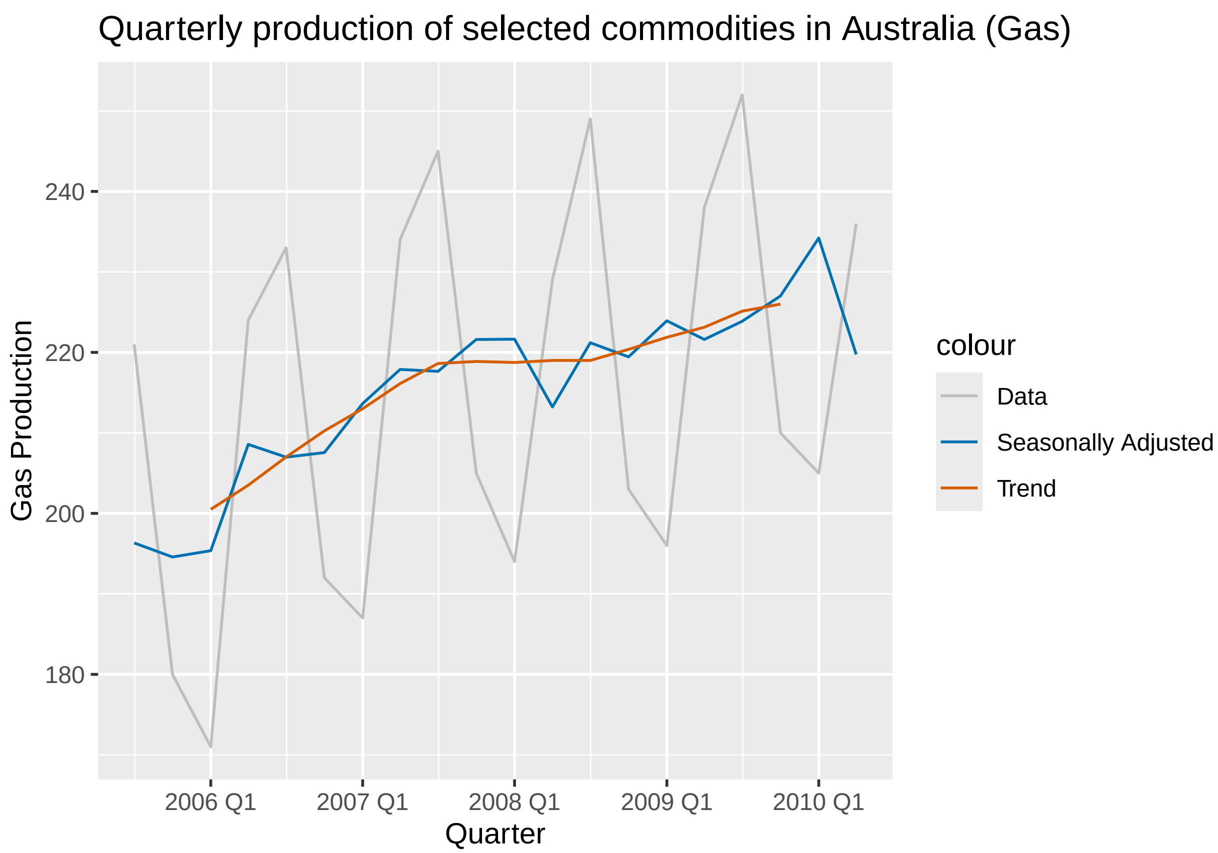
```
## Warning: Removed 2 rows containing missing values or values outside the scale range
## (`geom_line()`).
```



- c.Do the results support the graphical interpretation from part a?
- Answer: Yes, it seems every year's pattern are the same.
- d.Compute and plot the seasonally adjusted data.

```r
gas %>%
  model(classical_decomposition(Gas, type = "multiplicative")) %>%
  components() %>%
  ggplot(aes(x = Quarter)) +
  geom_line(aes(y = Gas, colour = "Data")) +
  geom_line(aes(y = season_adjust,
                colour = "Seasonally Adjusted")) +
  geom_line(aes(y = trend, colour = "Trend")) +
  labs(y = "Gas Production",
       title = "Quarterly production of selected commodities in Australia (Gas)") +
  scale_colour_manual(
    values = c("gray", "#0072B2", "#D55E00"),
    breaks = c("Data", "Seasonally Adjusted", "Trend")
  )
```

```
## Warning: Removed 4 rows containing missing values or values outside the scale range
## (`geom_line()`).
```



- e.Change one observation to be an outlier (e.g., add 300 to one observation), and recompute the seasonally adjusted data. What is the effect of the outlier?
- Answer: When I add 300 to the outlier, it changed the shape of the data plot, it create a outlier during 1st quarter of 2006.

```
gas2 <- gas
gas2$Gas[3] <- gas2$Gas[3] + 300
```

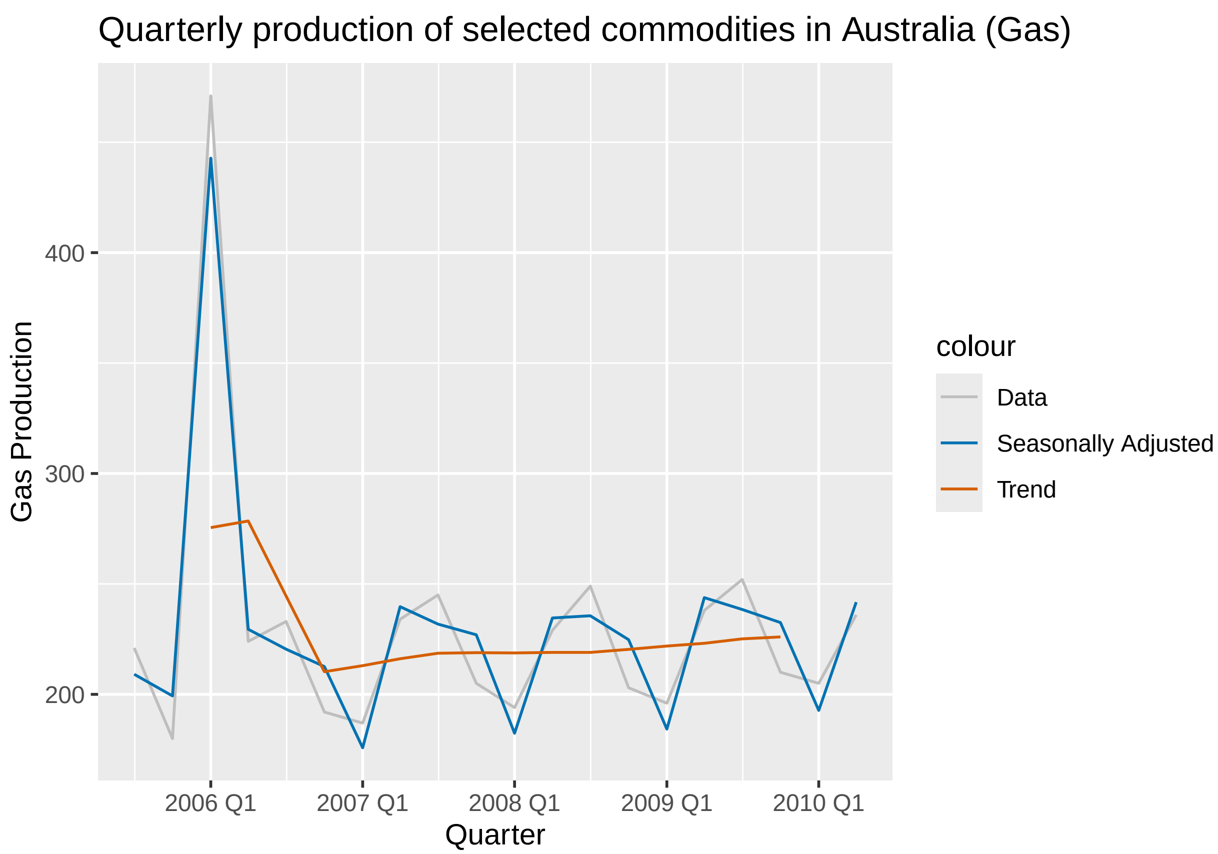
```
gas2
```

```
## # A tsibble: 20 x 2 [1Q]
##      Gas Quarter
##    <dbl>   <qtr>
##  1   221 2005 Q3
##  2   180 2005 Q4
##  3   471 2006 Q1
##  4   224 2006 Q2
##  5   233 2006 Q3
##  6   192 2006 Q4
##  7   187 2007 Q1
##  8   234 2007 Q2
##  9   245 2007 Q3
## 10   205 2007 Q4
## 11   194 2008 Q1
## 12   229 2008 Q2
```

```
## 13   249 2008 Q3
## 14   203 2008 Q4
## 15   196 2009 Q1
## 16   238 2009 Q2
## 17   252 2009 Q3
## 18   210 2009 Q4
## 19   205 2010 Q1
## 20   236 2010 Q2
```

```
gas2 %>%
  model(classical_decomposition(Gas, type = "multiplicative")) %>%
  components() %>%
  ggplot(aes(x = Quarter)) +
  geom_line(aes(y = Gas, colour = "Data")) +
  geom_line(aes(y = season_adjust,
                colour = "Seasonally Adjusted")) +
  geom_line(aes(y = trend, colour = "Trend")) +
  labs(y = "Gas Production",
       title = "Quarterly production of selected commodities in Australia (Gas)") +
  scale_colour_manual(
    values = c("gray", "#0072B2", "#D55E00"),
    breaks = c("Data", "Seasonally Adjusted", "Trend")
  )
```

```
## Warning: Removed 4 rows containing missing values or values outside the scale range
## (‘geom_line()‘).
```

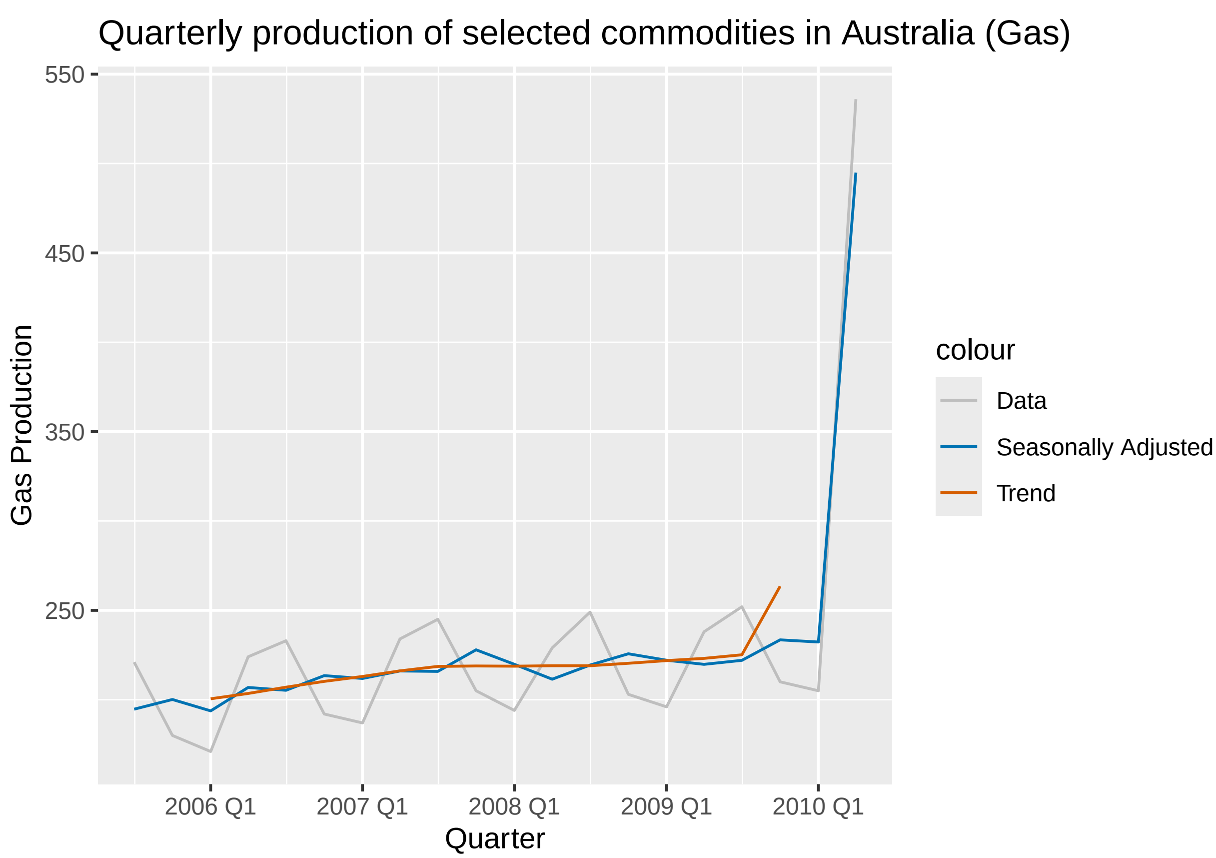


- f.Does it make any difference if the outlier is near the end rather than in the middle of the time series?
- Answer: It doesn't make any difference if the outlier is near the end rather than in the middle of the time series. If the outlier is present, it will change or adjusted the whole plot.

```
gas3 <- gas
gas3$Gas[20] <- gas2$Gas[20] + 300
```

```
gas3 %>%
  model(classical_decomposition(Gas, type = "multiplicative")) %>%
  components() %>%
  ggplot(aes(x = Quarter)) +
  geom_line(aes(y = Gas, colour = "Data")) +
  geom_line(aes(y = season_adjust,
                colour = "Seasonally Adjusted")) +
  geom_line(aes(y = trend, colour = "Trend")) +
  labs(y = "Gas Production",
       title = "Quarterly production of selected commodities in Australia (Gas)") +
  scale_colour_manual(
    values = c("gray", "#0072B2", "#D55E00"),
    breaks = c("Data", "Seasonally Adjusted", "Trend")
  )
```

```
## Warning: Removed 4 rows containing missing values or values outside the scale range
## (`geom_line()`).
```

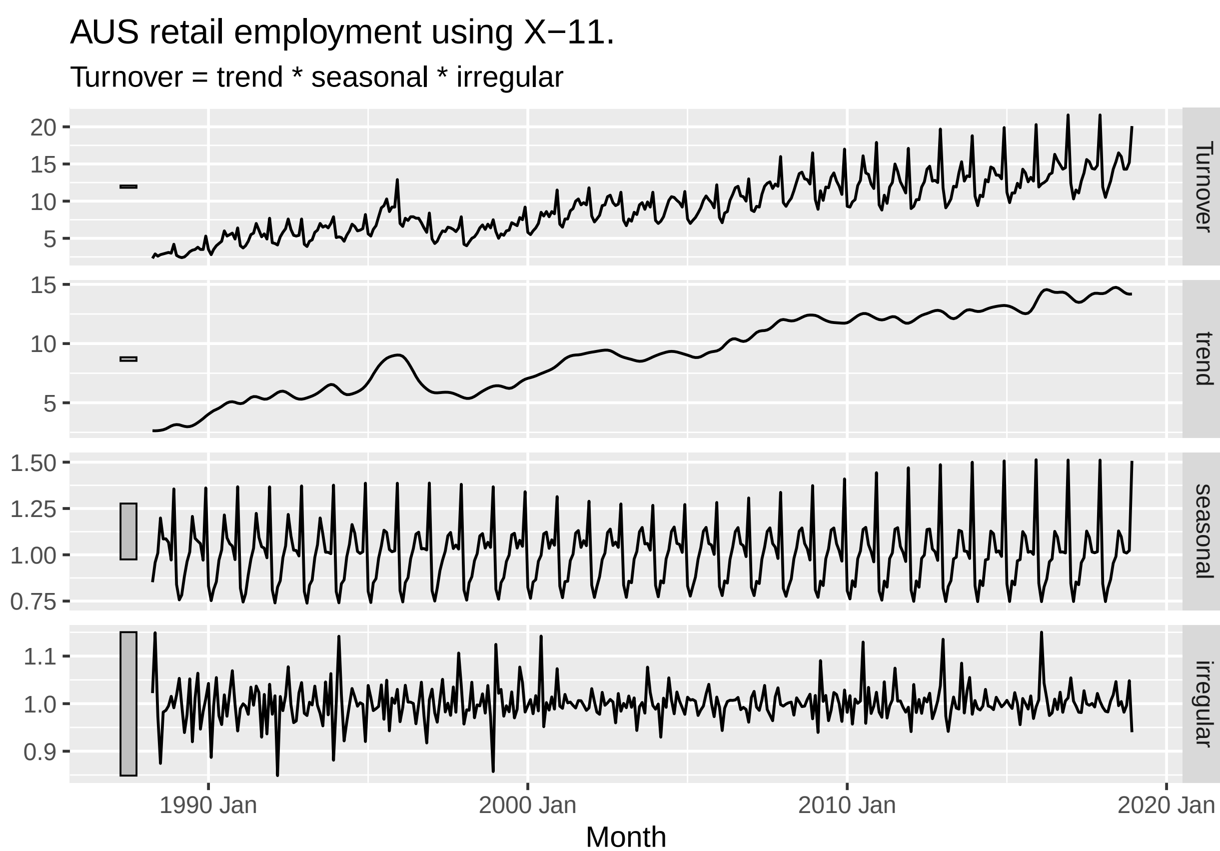


8.Recall your retail time series data (from Exercise 7 in Section 2.10). Decompose the series using X-11. Does it reveal any outliers, or unusual features that you had not noticed previously?

- Answer: Yes, in the plot, I see some peak trend, that may be the outliers. X-11 helps a lot to see the outliers in data.

```
set.seed(12345678)
myseries <- aus_retail %>%
  filter(`Series ID` == sample(aus_retail$`Series ID`,1))

x11_dcmp <- myseries |>
  model(x11 = X_13ARIMA_SEATS(Turnover ~ x11())) |>
  components()
autoplot(x11_dcmp) +
  labs(title =
    "AUS retail employment using X-11.")
```



9.Figures 3.19 and 3.20 show the result of decomposing the number of persons in the civilian labour force in Australia each month from February 1978 to August 1995. - a.Write about 3–5 sentences describing the results of the decomposition. Pay particular attention to the scales of the graphs in making your interpretation. - a Answer: The overall trend is going upward. I observed there are many sharp upticks and sharp declines. This pattern is repeated seasonally. And there are a couple of outliers during 1992 and 1998, the trend decrease a lot. - b.Is the recession of 1991/1992 visible in the estimated components? - b Answer:Yes