Chapter 3

Chi Hang(Philip) Cheung

2025-02-13

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(fpp3)
```

```
## 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 tidyr       1.3.1      v tsibbledata 0.4.1
## v lubridate   1.9.4      v feasts      0.4.1
## v ggplot2     3.5.1      v fable       0.4.1
```

```
## -- 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)
```

```
##
## 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?

Ans: Monaco has the highest GDP per capita. The overall trend for all countries are increasing.

```
per_cap <- global_economy %>%
  group_by(Country) %>%
  mutate(percap = GDP/Population) %>%
  select(Country, Year, percap)


ggplot(per_cap, aes(x=Year, y=percap, color = Country)) +
  geom_line() +
  theme(legend.position = 'none') +
  labs(title = 'All countries per capita', y='$USD')
```
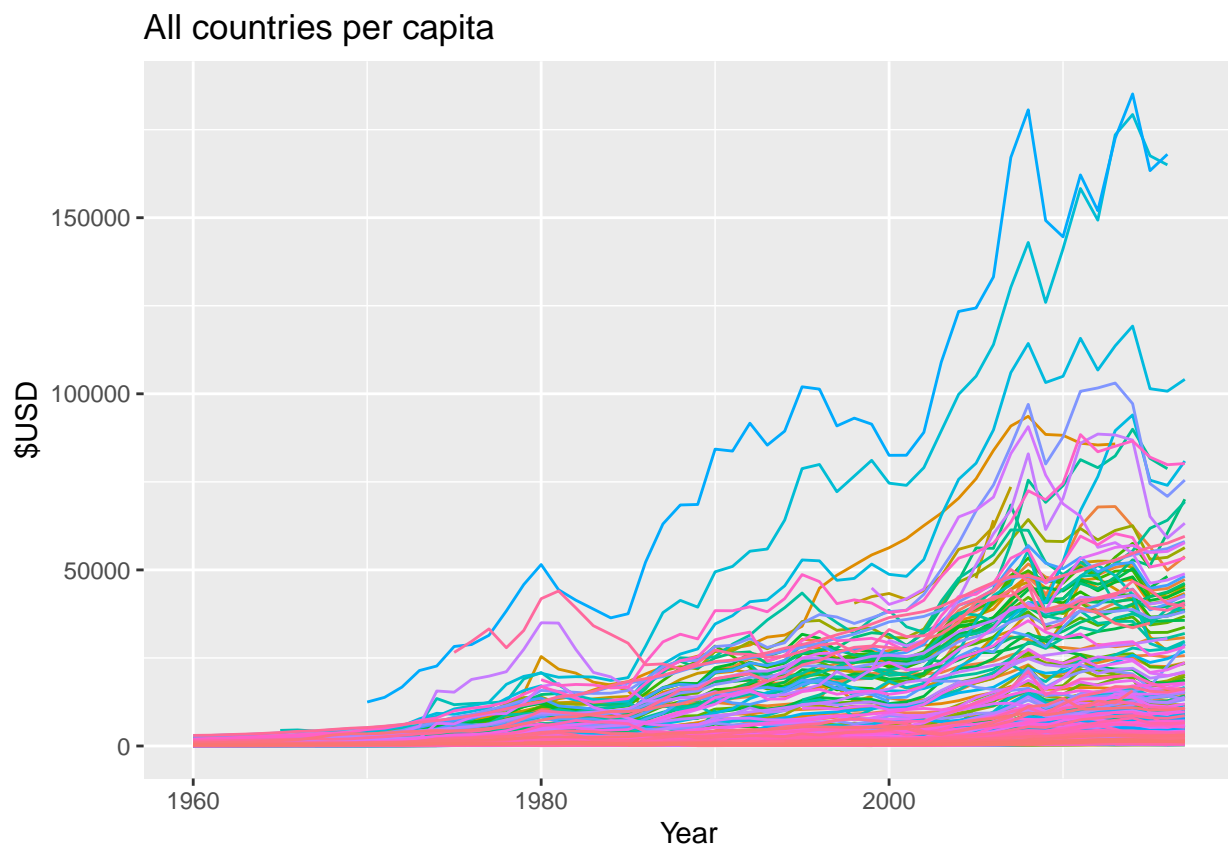

The above plot has too much data to display the legend. So i used another code below to extract the country with the highest GDP per capita. And that country is Monanco in year 2014

```
global_economy %>%
  group_by(Country) %>%
  summarise(per_cap_avg = mean(GDP / Population, na.rm = TRUE)) %>%
  slice_max(per_cap_avg)
```

```
## # A tsibble: 1 x 3 [1Y]
## # Key:        Country [1]
##   Country  Year per_cap_avg
##   <fct>   <dbl>       <dbl>
## 1 Monaco   2014     185153.
```

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

- United States GDP from `global_economy`.
- Slaughter of Victorian "Bulls, bullocks and steers" in `aus_livestock`.
- Victorian Electricity Demand from `vic_elec`.
- Gas production from `aus_production`.

US GDP

Ans: There is not much difference between the adjusted and unadjusted data. The main difference is seen in the scaling of y-axis–a smaller scale is seen when adjusted to population. USA has been on the rise.

```
#global_economy for USA:
global_economy %>%
  filter(Code == 'USA') %>%
  index_by(Year) %>%
  mutate(GDP_per_capita = GDP/Population) %>%
  pivot_longer(c(GDP, GDP_per_capita),
               values_to='wealth',
               names_to = 'GDP_type') %>%
  ggplot(aes(x=Year, y=wealth)) +
  geom_line()+
  facet_grid(GDP_type~., scales = 'free_y') +
  labs(title='USA GDP and GDP/capita', y='USD$')
```

USA GDP and GDP/capita
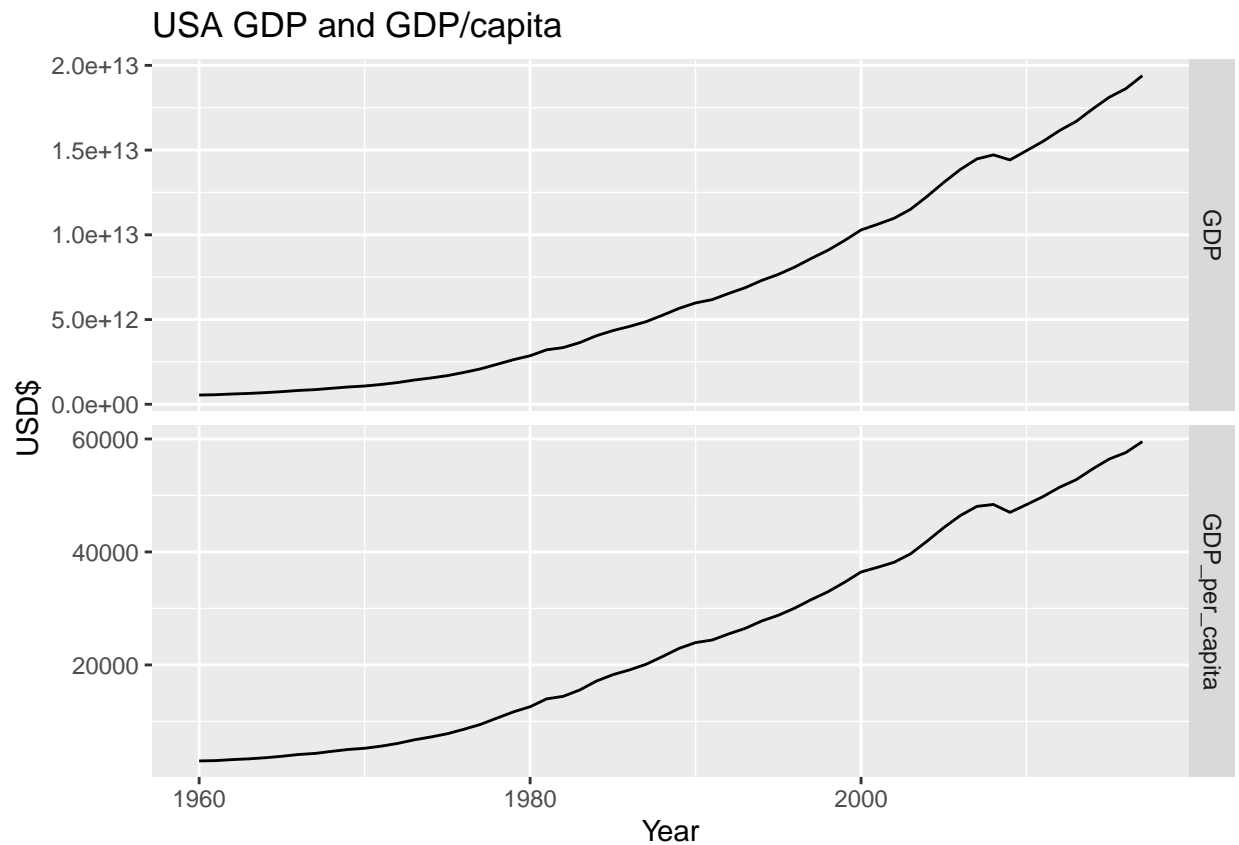
Australia slaughter house data

Ans: The transformation changes the scale and slightly stabilized the variance of the data. There seems to be a seasonality with a cyclic pattern trend in the data.

```r
library(latex2exp)
livestock_og<- aus_livestock %>%
  filter(Animal == 'Bulls, bullocks and steers', State=='Victoria')

livestock_lambda<- aus_livestock %>%
  filter(Animal == 'Bulls, bullocks and steers', State=='Victoria') %>%
  features(Count, features = guerrero) %>%
  pull(lambda_guerrero)

livestock_og %>%
  autoplot(Count)+
  labs(title = 'Victoria State Bulls, bullocks, and steers slaugther number',
       y='numbers slaugthered')
```
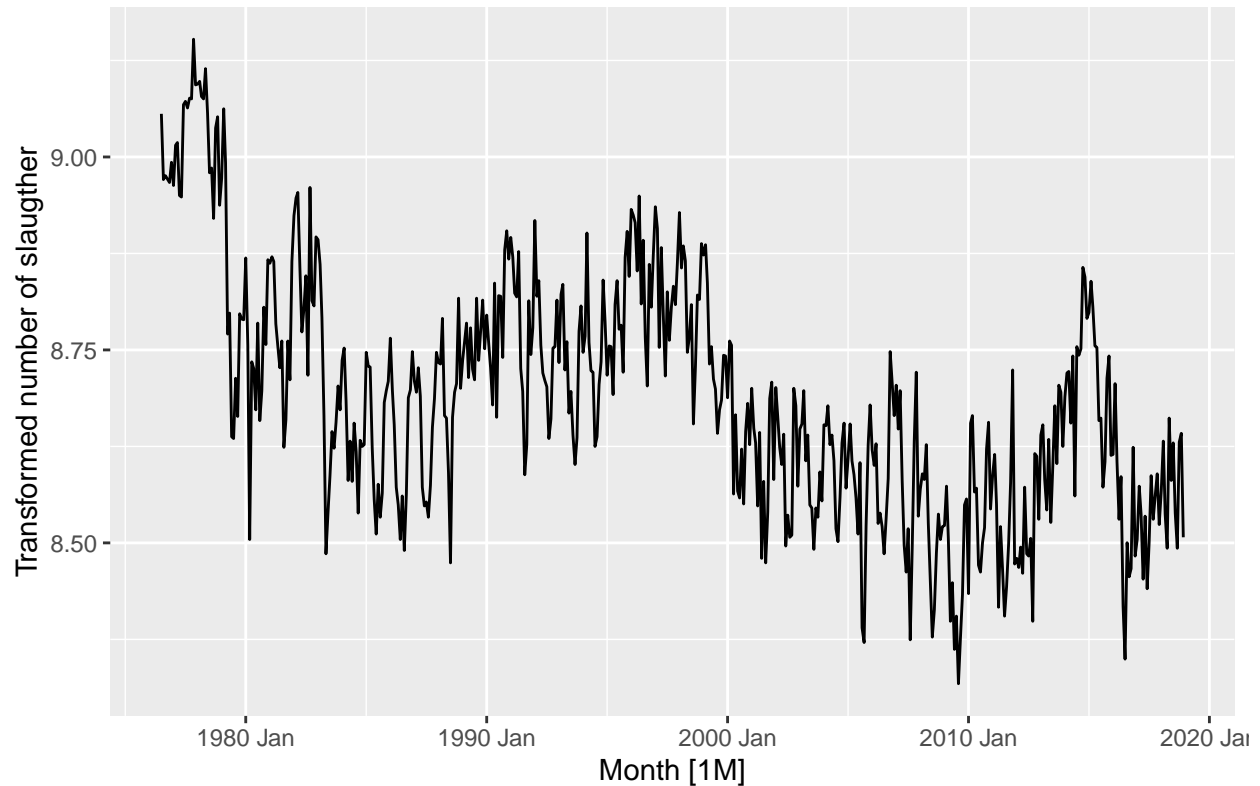
Victoria State Bulls, bullocks, and steers slaugther number
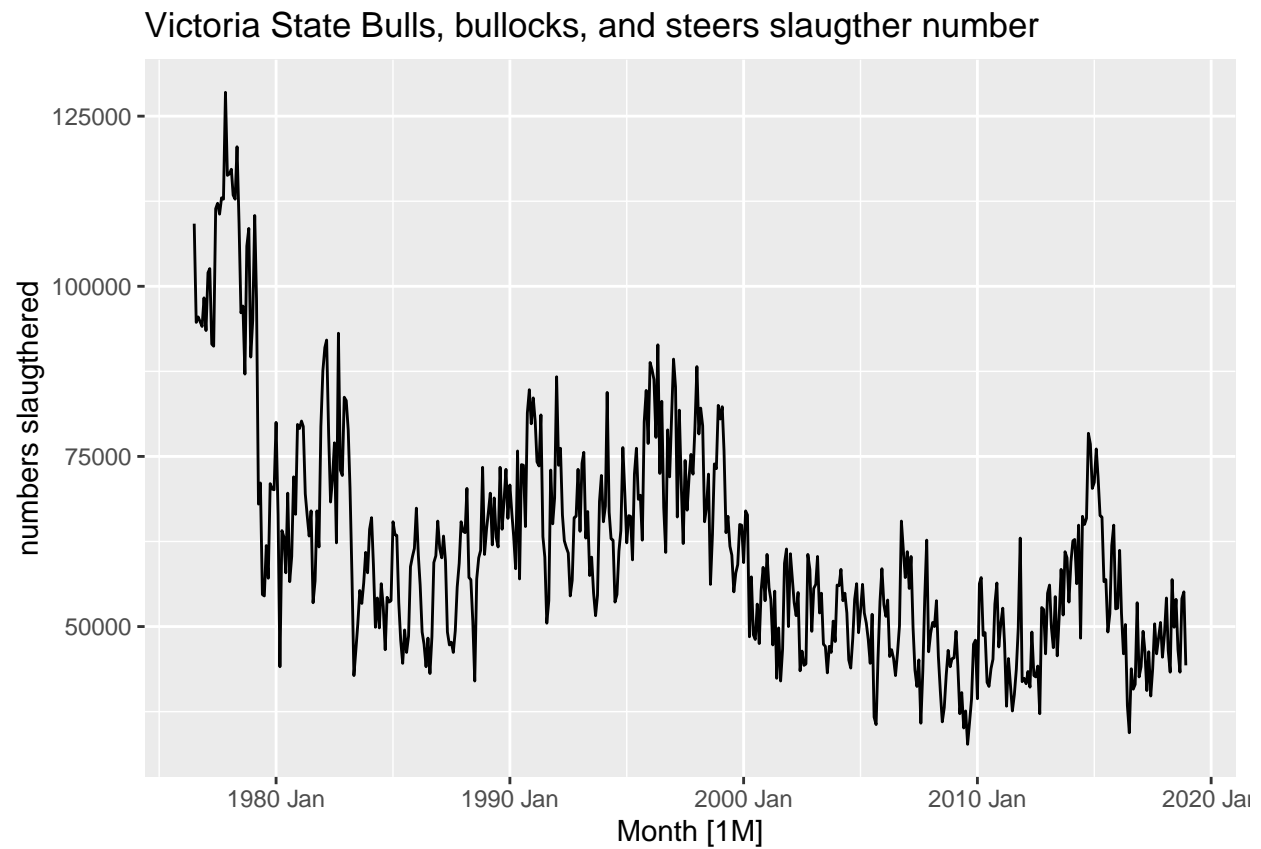

```r
livestock_og %>%
  autoplot(box_cox(Count, livestock_lambda)) +
  labs(title = latex2exp::TeX(paste0('Victoria State Bulls, bullocks, and steers slaugther number. lamb
      y='Transformed number of slaugther')
```
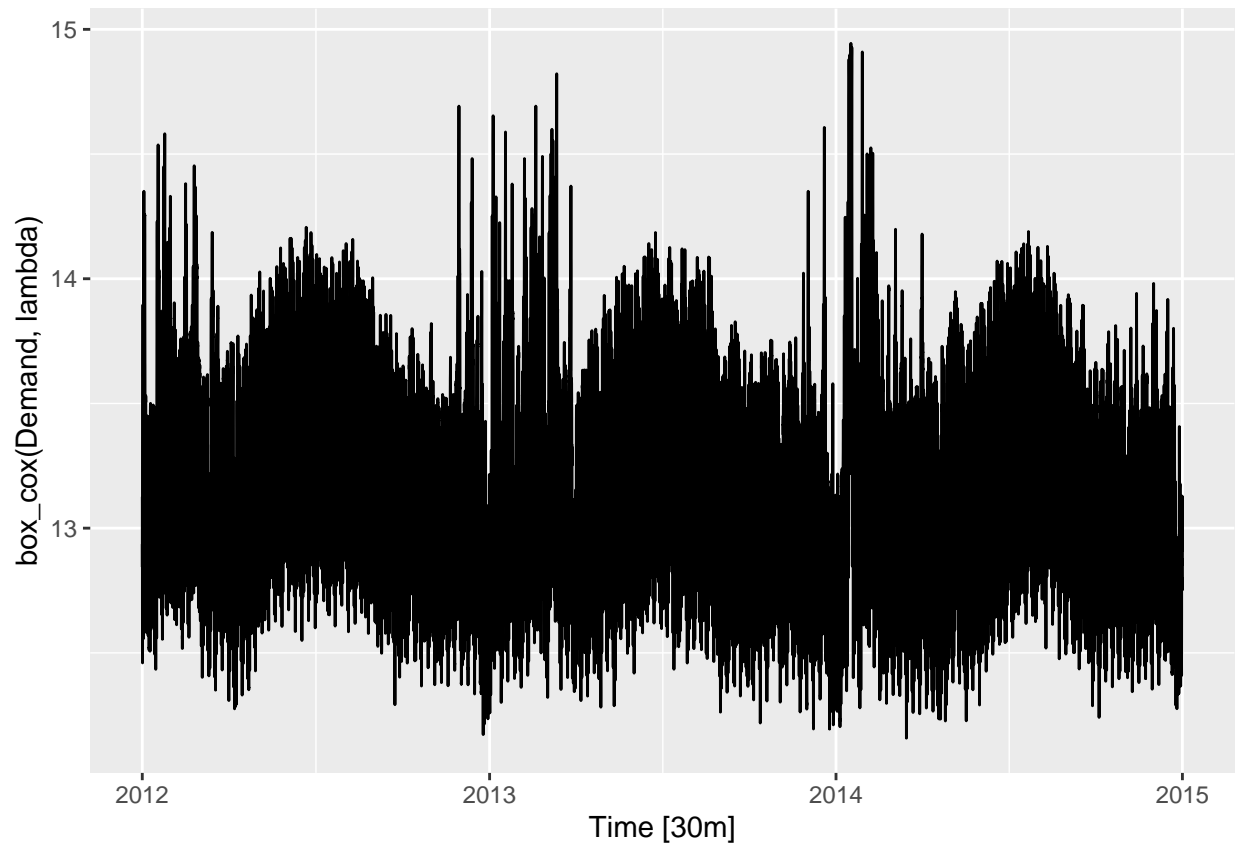
Victoria State Bulls, bullocks, and steers slaugther number. lambda =−0.04

Victorian electricity demand data

Ans: The transformation did not yield much difference. This could be that the variance from the original data is already relatively even.

```r
lambda <- vic_elec %>%
  features(Demand, features = guerrero) %>%
  pull(lambda_guerrero)

vic_elec %>%
  autoplot(box_cox(Demand, lambda))
```
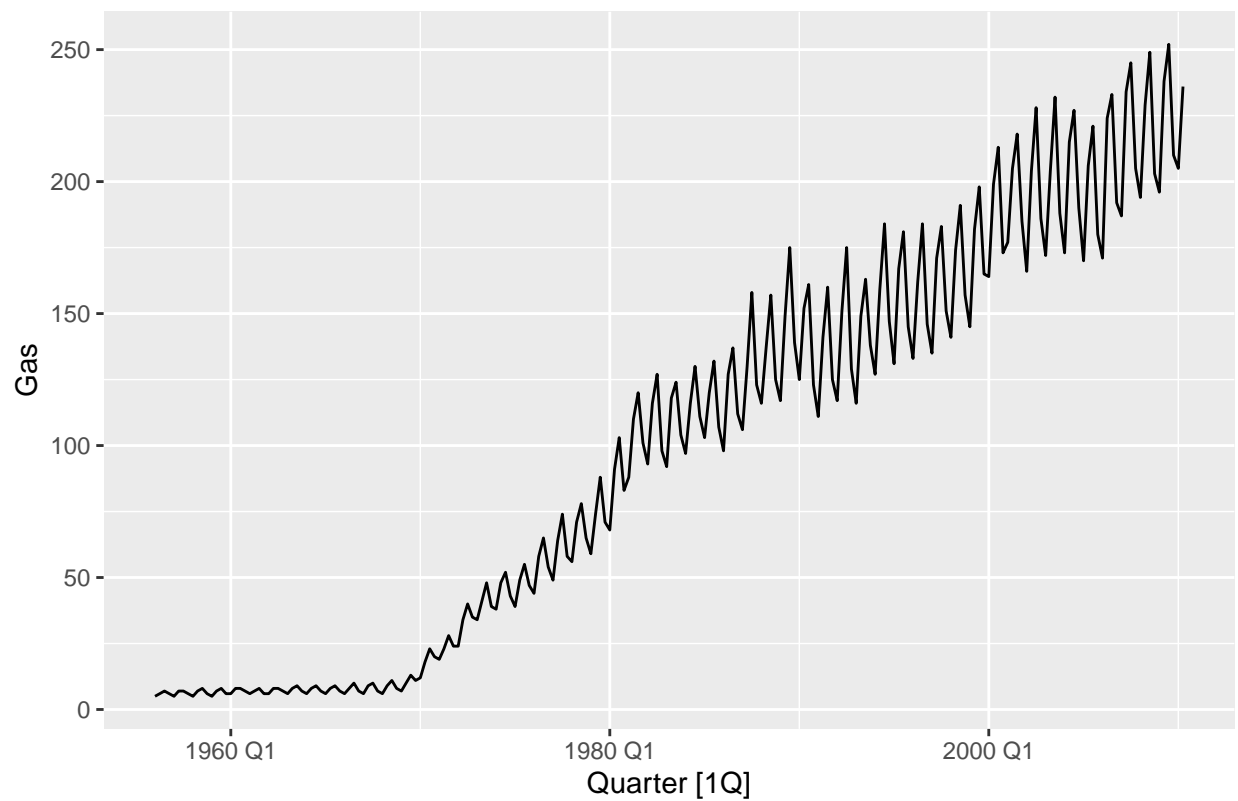
Gas production

Ans: The box-cox transformation helped to create a more even variance amount the data when compared to the original data, where the start of the data are not that pronounced.

```
gas<- aus_production %>%
  select(Quarter, Gas)

gas_lambda<- aus_production %>%
  select(Quarter, Gas) %>%
  features(Gas, features = guerrero) %>%
  pull(lambda_guerrero)

gas %>%
  autoplot(Gas)+
  labs(title = 'AUS Quarterly Gas production')
```
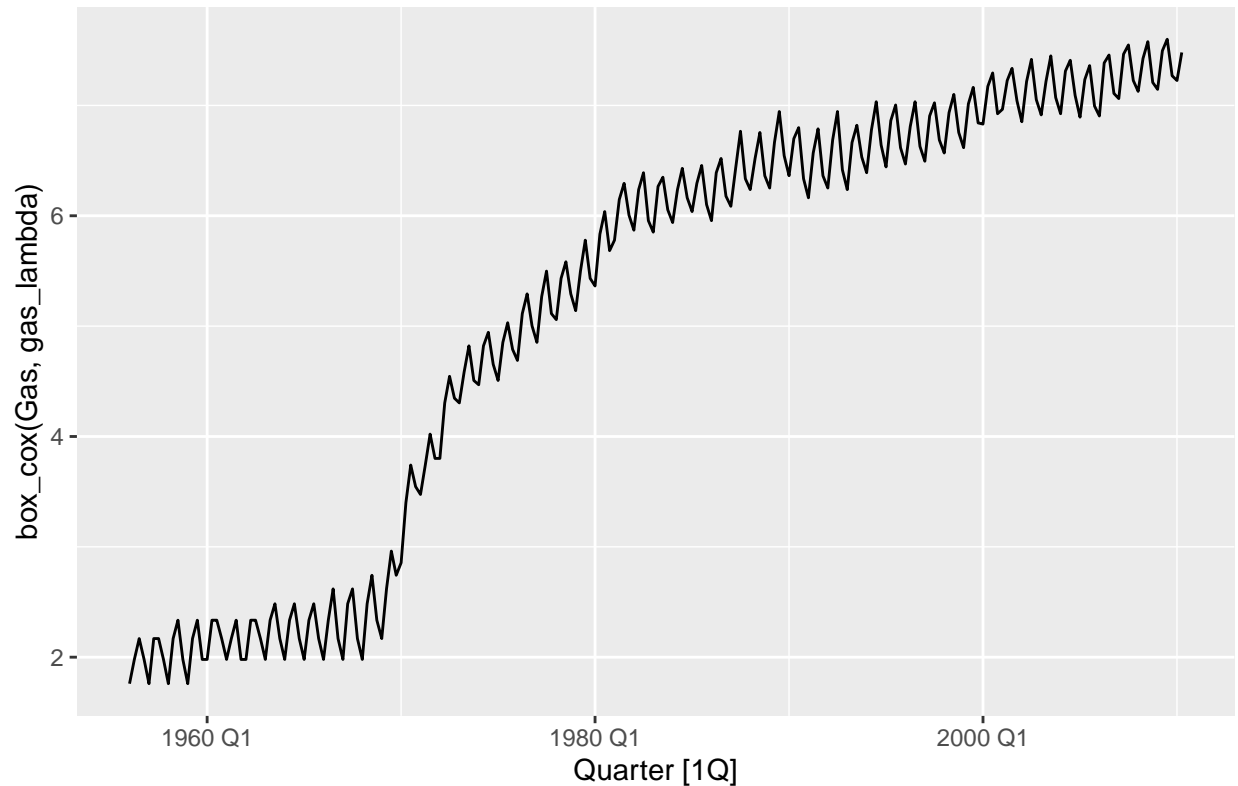
AUS Quarterly Gas production


```r
gas %>%
  autoplot(box_cox(Gas, gas_lambda))+
  labs(title = latex2exp::TeX(paste0('Transformed AUS Quarterly Gas production. Lambda =', round(gas_lar
```

## Transformed AUS Quarterly Gas production. Lambda =0.11Transformed gas

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

   Ans: This could be that the original data already has a well spread of variance and further transformation will yield similar graph.
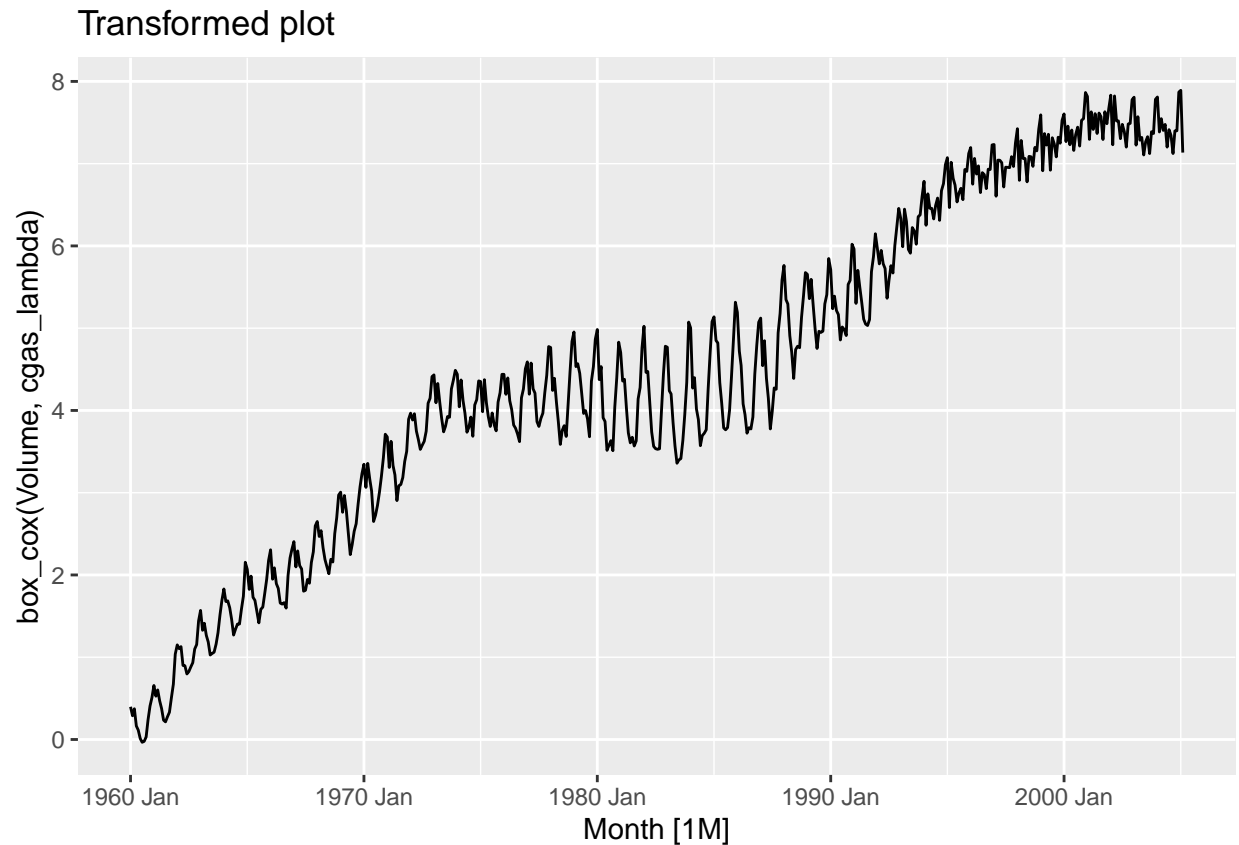
```
cgas_lambda<- canadian_gas %>%
  features(Volume, features = guerrero) %>%
  pull(lambda_guerrero)

canadian_gas %>%
  autoplot()+labs(title = 'original plot')
```

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

original plot
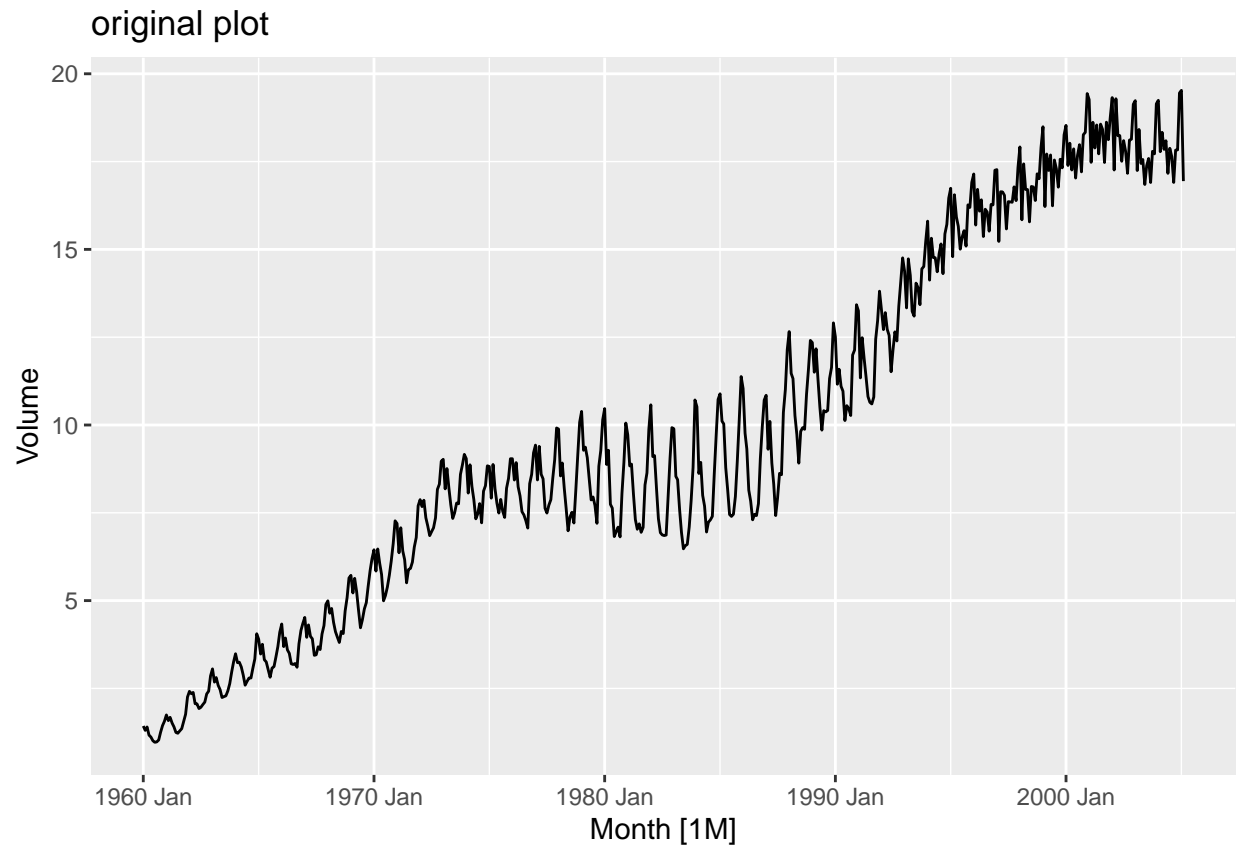

```
canadian_gas %>%
  autoplot(box_cox(Volume, cgas_lambda))+
  labs(title = 'Transformed plot')
```

Transformed plot

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

   Ans: Lambda of 0.215 is likely the most optimal Box Cox lambda value.

```
set.seed(123)
myseries <- aus_retail |>
  filter(`Series ID` == sample(aus_retail$`Series ID`,1)) %>%
  features(Turnover, features = guerrero) %>%
  pull(lambda_guerrero)

myseries
```

```
## [1] 0.2151641
```

5. For the following series, find an appropriate Box-Cox transformation in order to stabilize the variance.

   a. Tobacco from aus_production

Ans: Lambda = 0.926

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

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

b) Economy class passengers between Melbourne and Sydney from `ansett`

Ans: lambda = 1.99

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

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

c) Pedestrian counts at Southern Cross Station from pedestrian.

Ans: lambda = -0.25

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

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

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?

Ans: The decomposition reveals that the seasonality is slowly shifting to a lesser turnover pattern as time goes by despite the overall growing trend. There seems to be a potential outlier highlighted a little pass January of 2000, where an unusual peak is observed in the remainder series, but the scale of it is very small.

```
set.seed(123)
x11_dcmp<- aus_retail |>
  filter(`Series ID` == sample(aus_retail$`Series ID`,1)) %>%
  model(x11 = X_13ARIMA_SEATS(Turnover ~ x11())) %>%
  components()

autoplot(x11_dcmp) +
  labs(title = 'Aus Retail decomposition using x11')
```

Aus Retail decomposition using x11
Turnover = trend * seasonal * irregular
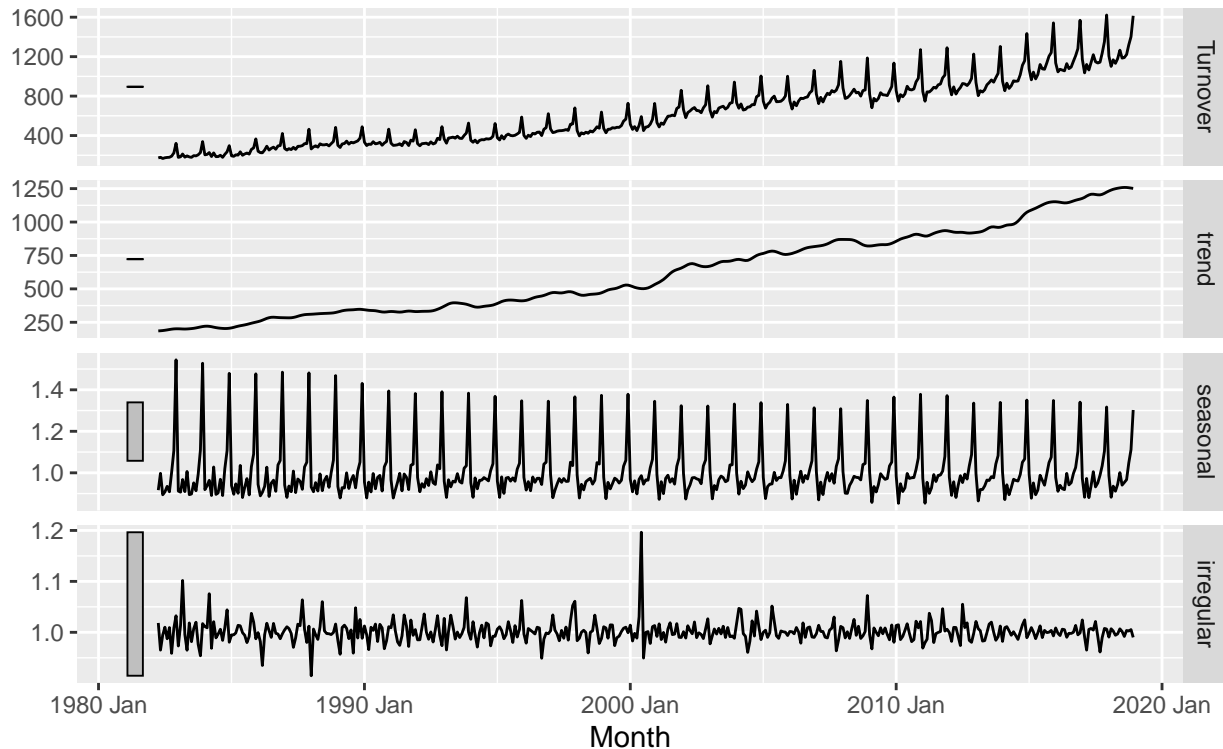
9. Figures 3.19 and 3.20 show the result of decomposing the number of persons in the civilian labour force in Australia each month from February 1978 to August 1995.

   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.

   Ans: There is a general growth of labor force as the year went on. The seasonality decomposition reveals that around December the demand for labor force has slightly increased over time. Similarly, the trough of labor force, around January is also dipped more when compared to previous years. Seasonality, however, as shown by the scale on the left side, is the smallest factor amount the entire decomposition series.

   b) Is the recession of 1991/1992 visible in the estimated components?

   Ans: The recession of 1991/1992 is very visible as shown in the remainder series, where a sudden decline is observed.