Homework 2

Aaron Banlao

```r
library(fpp3)
```

```
## -- Attaching packages -------------------------------------------- fpp3 0.4.0 --

## v tibble      3.1.8          v tsibble     1.1.3
## v dplyr       1.0.10         v tsibbledata 0.4.1.9000
## v tidyr       1.2.1          v feasts      0.3.0
## v lubridate   1.9.0          v fable       0.3.2
## v ggplot2     3.4.0

## -- 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(tsibbledata)
library(seasonal)
```

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

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

Problem 7

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

gas <- tail(aus_production, 5*4) |> select(Gas) Plot the time series. Can you identify seasonal fluctuations and/or a trend-cycle? Use classical_decomposition with type=multiplicative to calculate the trend-cycle and seasonal indices. Do the results support the graphical interpretation from part a? Compute and plot the seasonally adjusted data. 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? Does it make any difference if the outlier is near the end rather than in the middle of the time series?

a)

Answer)

There seems to be a decline at every year's Q1 as well as a peak in every year's Q2. There are definitely seasonal fluctuations with a slight upward trend

Code)

```
head(aus_production)
```

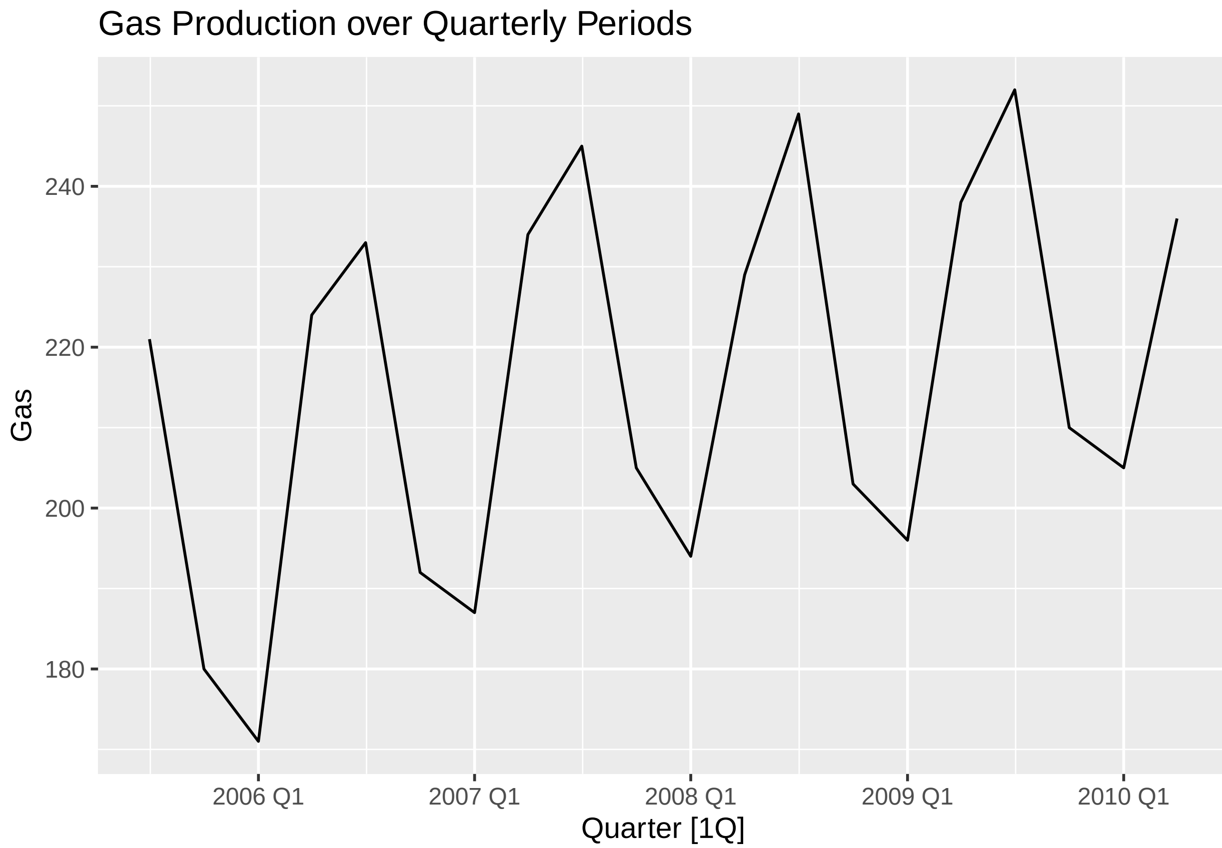
```
## # A tsibble: 6 x 7 [1Q]
##   Quarter  Beer Tobacco Bricks Cement Electricity   Gas
##     <qtr> <dbl>   <dbl>  <dbl>  <dbl>       <dbl> <dbl>
## 1 1956 Q1   284    5225    189    465        3923     5
## 2 1956 Q2   213    5178    204    532        4436     6
## 3 1956 Q3   227    5297    208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

```
gas <- tail(aus_production, 5*4) %>%
  select(Gas)
head(gas)
```

```
## # A tsibble: 6 x 2 [1Q]
##     Gas Quarter
##   <dbl>   <qtr>
## 1   221 2005 Q3
## 2   180 2005 Q4
## 3   171 2006 Q1
## 4   224 2006 Q2
## 5   233 2006 Q3
## 6   192 2006 Q4
```

```
autoplot(gas) + labs(title = "Gas Production over Quarterly Periods")
```

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

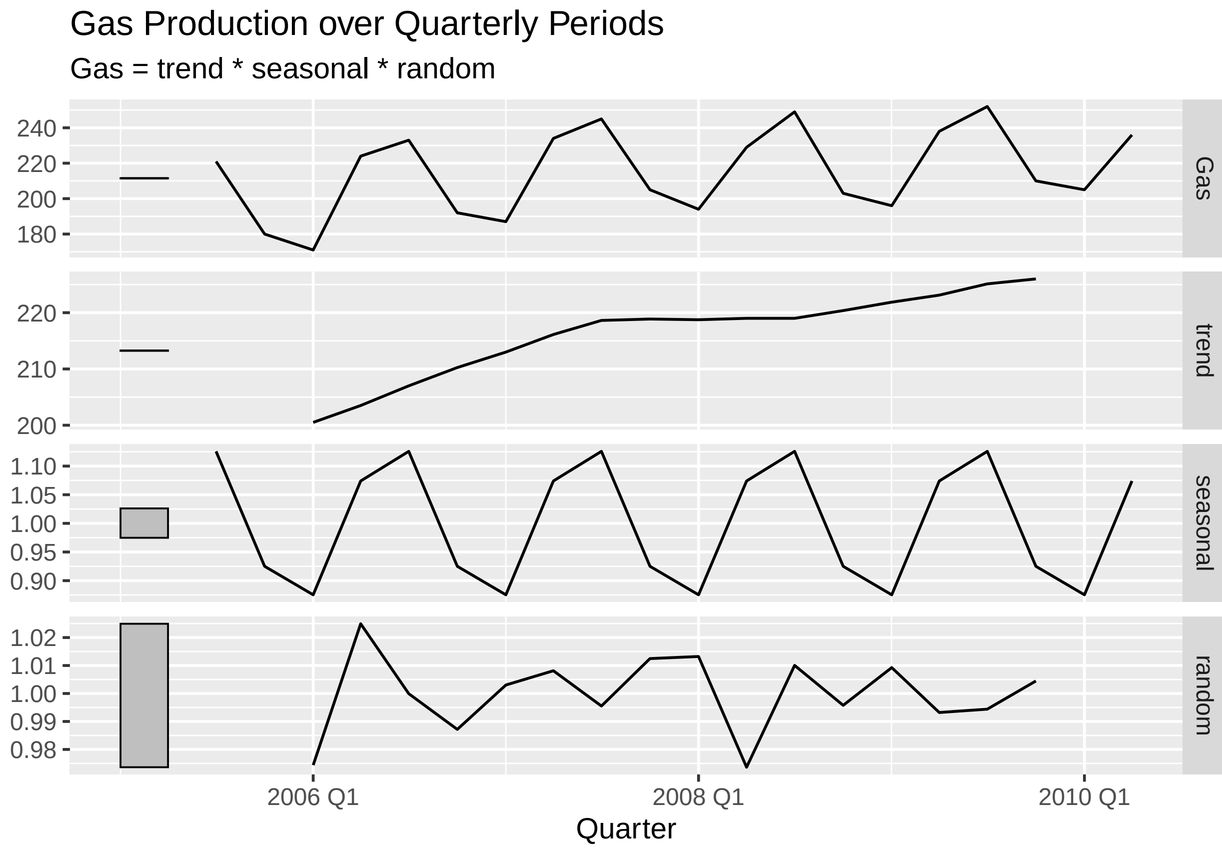


b)

Code)

```
gas %>%
  model(
    classical_decomposition(Gas, type = "multiplicative")
  ) %>%
  components() %>%
  autoplot() + labs(title = "Gas Production over Quarterly Periods")
```

```
## Warning: Removed 2 rows containing missing values (‘geom_line()‘).
```



c)

Answer)

Yes, the results support our assumptions from question a. There is an apparent trend and there is a clear seasonal pattern according to the plot.

d)

Code)

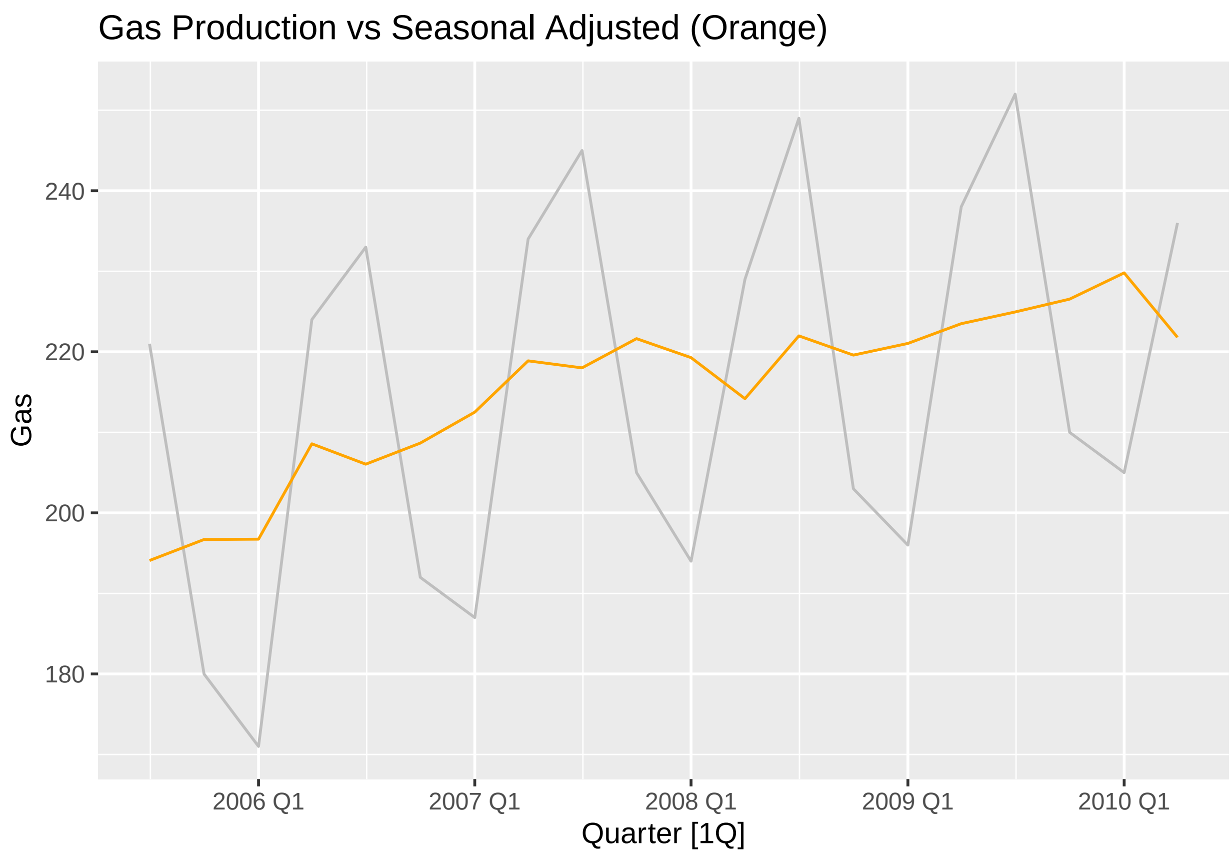
```
stlgas <- gas %>%
  model(stl = STL(Gas))

components(stlgas)
```

```
## # A dable: 20 x 7 [1Q]
## # Key:     .model [1]
## # :        Gas = trend + season_year + remainder
##    .model Quarter   Gas trend season_year remainder season_adjust
##    <chr>    <qtr> <dbl> <dbl>       <dbl>     <dbl>         <dbl>
##  1 stl    2005 Q3   221  193.        26.9     0.856          194.
```

```
##  2 stl    2005 Q4   180  197.        -16.7    -0.109           197.
##  3 stl    2006 Q1   171  200.        -25.7    -3.59            197.
##  4 stl    2006 Q2   224  204.         15.4     4.86            209.
##  5 stl    2006 Q3   233  207.         27.0    -1.03            206.
##  6 stl    2006 Q4   192  210.        -16.7    -1.33            209.
##  7 stl    2007 Q1   187  213.        -25.5    -0.550           213.
##  8 stl    2007 Q2   234  216.         15.1     2.60            219.
##  9 stl    2007 Q3   245  219.         27.0    -0.730           218.
## 10 stl    2007 Q4   205  219.        -16.6     2.55            222.
## 11 stl    2008 Q1   194  219.        -25.3     0.562           219.
## 12 stl    2008 Q2   229  219.         14.8    -4.59            214.
## 13 stl    2008 Q3   249  219.         27.0     2.98            222.
## 14 stl    2008 Q4   203  220.        -16.6    -0.834           220.
## 15 stl    2009 Q1   196  222.        -25.0    -0.740           221.
## 16 stl    2009 Q2   238  223.         14.5     0.341           223.
## 17 stl    2009 Q3   252  225.         27.1    -0.132           225.
## 18 stl    2009 Q4   210  226.        -16.5     0.986           227.
## 19 stl    2010 Q1   205  226.        -24.8     4.25            230.
## 20 stl    2010 Q2   236  225.         14.2    -3.62            222.
```

```
components(stlgas) %>%
  as_tsibble() %>%
  autoplot(Gas, colour= "gray") +
  geom_line(aes(y = season_adjust), color = "orange") + labs(title = "Gas Production vs Seasonal Adjust
```



e)

Answer)

The outlier changed substantially changed the shape of the data plot

Code)

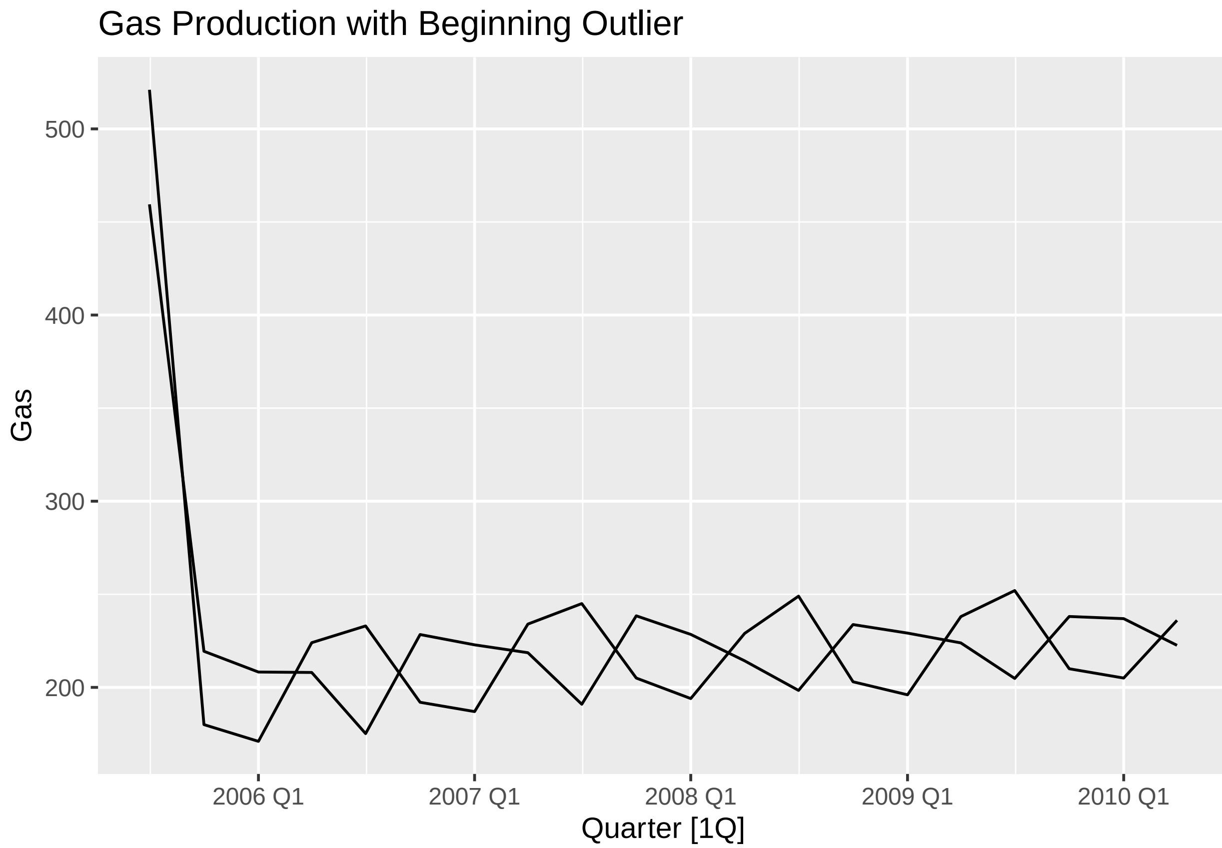
```
gas2 <- gas
```

```
gas2$Gas[1] <- gas2$Gas[1] + 300
```

```
stlgas2 <- gas2 %>%
  model(stl = STL(Gas))
components(stlgas2)
```

```
## # A dable: 20 x 7 [1Q]
## # Key:     .model [1]
## # :        Gas = trend + season_year + remainder
##    .model Quarter   Gas trend season_year remainder season_adjust
##    <chr>    <qtr> <dbl> <dbl>       <dbl>     <dbl>         <dbl>
##  1 stl    2005 Q3   521  373.        61.5     86.6           459.
##  2 stl    2005 Q4   180  310.       -39.4    -90.1           219.
##  3 stl    2006 Q1   171  250.       -37.2    -42.0           208.
##  4 stl    2006 Q2   224  203.        16.0      5.18          208.
##  5 stl    2006 Q3   233  205.        57.8    -30.1           175.
##  6 stl    2006 Q4   192  210.       -36.4     18.1           228.
##  7 stl    2007 Q1   187  215.       -35.8      7.62          223.
##  8 stl    2007 Q2   234  216.        15.4      3.06          219.
##  9 stl    2007 Q3   245  217.        54.1    -26.2           191.
## 10 stl    2007 Q4   205  219.       -33.4     19.1           238.
## 11 stl    2008 Q1   194  221.       -34.5      7.85          228.
## 12 stl    2008 Q2   229  218.        14.7     -3.94          214.
## 13 stl    2008 Q3   249  218.        50.6    -19.2           198.
## 14 stl    2008 Q4   203  221.       -30.7     13.2           234.
## 15 stl    2009 Q1   196  223.       -33.2      5.69          229.
## 16 stl    2009 Q2   238  223.        14.1      1.14          224.
## 17 stl    2009 Q3   252  224.        47.2    -19.1           205.
## 18 stl    2009 Q4   210  226.       -28.1     11.9           238.
## 19 stl    2010 Q1   205  228.       -31.9      8.49          237.
## 20 stl    2010 Q2   236  231.        13.4     -8.31          223.
```

```
components(stlgas2) %>%
  as_tsibble() %>%
  autoplot(Gas) + labs(title = "Gas Production with Beginning Outlier") +
  geom_line(aes(y = season_adjust))
```



f)

Answer)

It doesn't seem to matter when the outlier is in the beginning or the end of the time series. The same shape is produced when plotting the data

Code)

```
gas3 <- gas
```

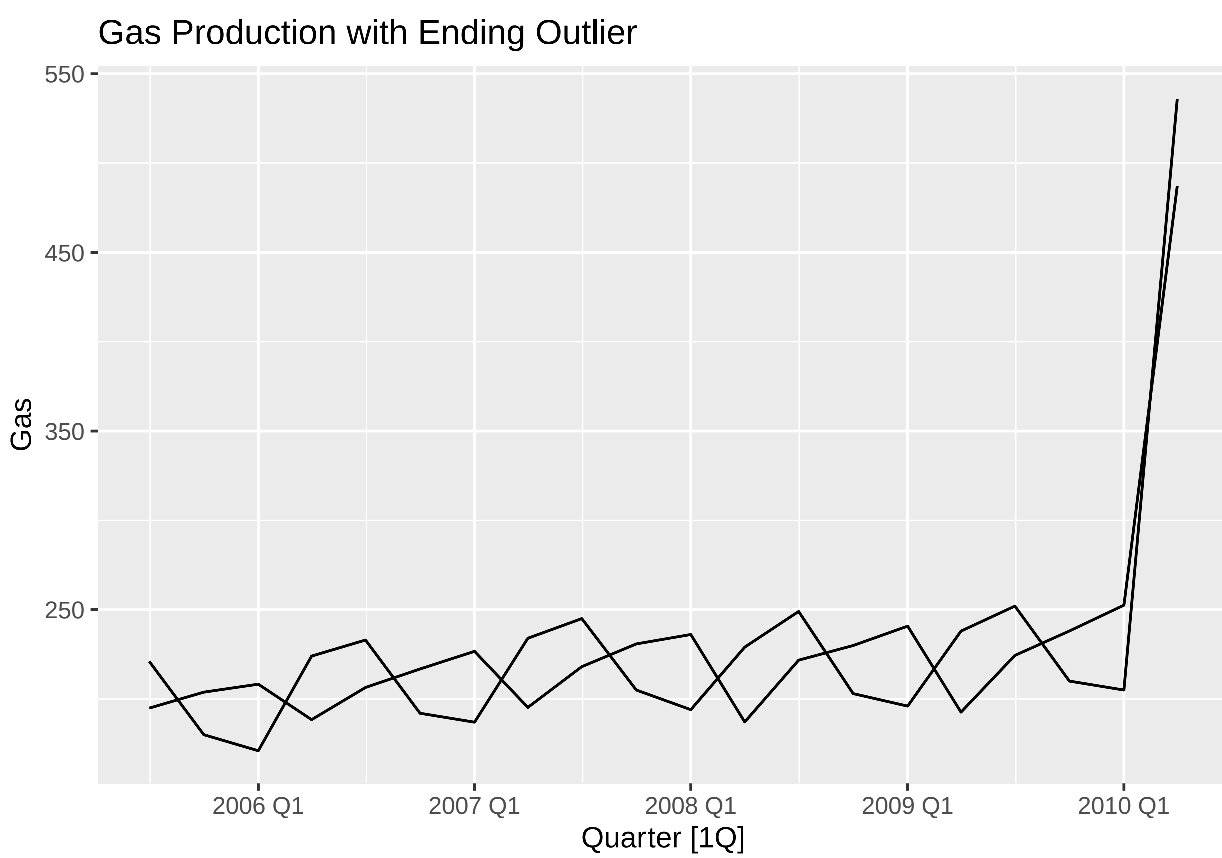
```
gas3$Gas[20] <- gas3$Gas[20] + 300
```

```
stlgas3 <- gas3 %>%
  model(stl = STL(Gas))
components(stlgas3)
```

```
## # A dable: 20 x 7 [1Q]
## # Key:     .model [1]
## # :        Gas = trend + season_year + remainder
##    .model Quarter   Gas trend season_year remainder season_adjust
##    <chr>    <qtr> <dbl> <dbl>       <dbl>     <dbl>         <dbl>
##  1 stl    2005 Q3   221  199.        26.1     -3.83          195.
```

```
## 2 stl     2005 Q4   180  200.        -23.8    4.13            204.
## 3 stl     2006 Q1   171  201.        -37.3    7.29            208.
## 4 stl     2006 Q2   224  203.         35.6  -14.1             188.
## 5 stl     2006 Q3   233  207.         26.6   -0.235           206.
## 6 stl     2006 Q4   192  212.        -24.8    5.10            217.
## 7 stl     2007 Q1   187  213.        -39.7   13.5             227.
## 8 stl     2007 Q2   234  215.         38.7  -19.6             195.
## 9 stl     2007 Q3   245  218.         26.9   -0.0731          218.
## 10 stl    2007 Q4   205  221.        -25.8    9.84            231.
## 11 stl    2008 Q1   194  219.        -42.1   17.1             236.
## 12 stl    2008 Q2   229  217.         41.9  -30.1             187.
## 13 stl    2008 Q3   249  218.         27.3    3.43            222.
## 14 stl    2008 Q4   203  223.        -26.9    7.34            230.
## 15 stl    2009 Q1   196  222.        -44.8   18.7             241.
## 16 stl    2009 Q2   238  221.         45.4  -28.8             193.
## 17 stl    2009 Q3   252  224.         27.6    0.184           224.
## 18 stl    2009 Q4   210  275.        -28.1  -37.4             238.
## 19 stl    2010 Q1   205  338.        -47.5  -85.7             253.
## 20 stl    2010 Q2   536  405.         48.8   82.2             487.
```

```
components(stlgas3) %>%
  as_tsibble() %>%
  autoplot(Gas) + labs(title = "Gas Production with Ending Outlier") +
  geom_line(aes(y = season_adjust))
```



Problem 8)

Recall your retail time series data (from Exercise 8 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)

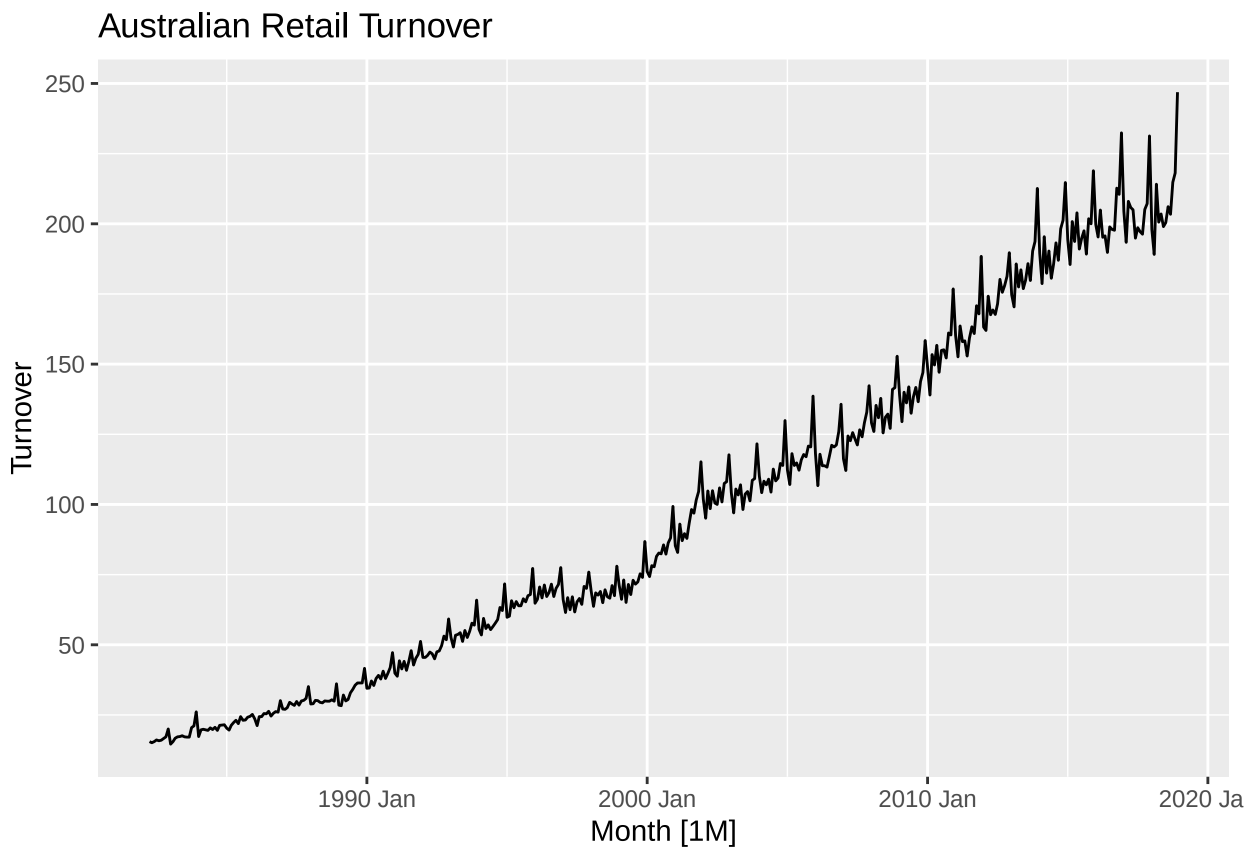
There are several spikes that occur in the remainder around the years 1982-1983. It is surprising because the seasonality and trend look consistent while visualizing turnover over time.

Code)

```
set.seed(101)
myseries <- aus_retail %>%
  filter(`Series ID` == sample(aus_retail$`Series ID`,1))
```

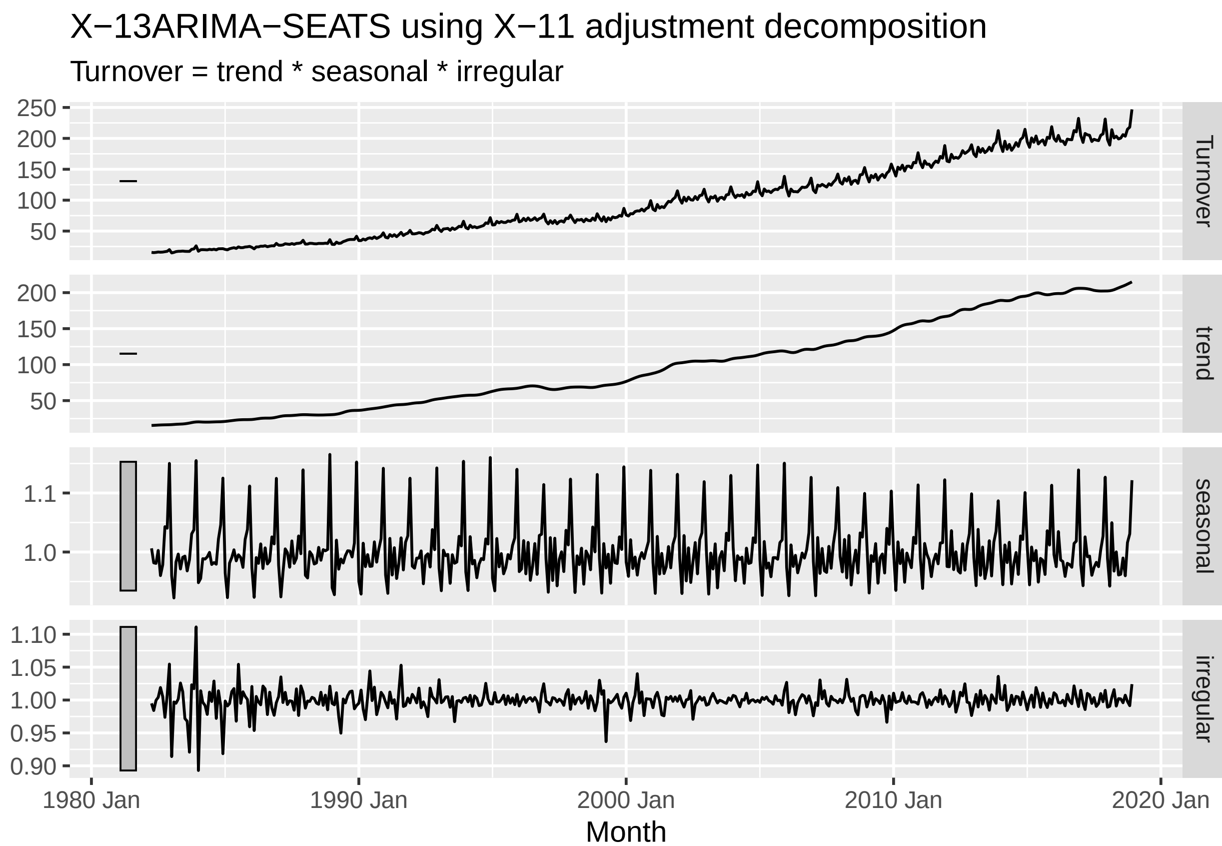
```
autoplot(myseries) + labs(title = "Australian Retail Turnover")
```

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



```r
x11myseries <- myseries %>%
  model(x11 = X_13ARIMA_SEATS(Turnover ~ x11())) %>%
  components()

autoplot(x11myseries)
```



Problem 9)

Write about 3–5 sentences describing the results of the decomposition. Pay particular attention to the scales of the graphs in making your interpretation. Is the recession of 1991/1992 visible in the estimated components?

Answer)

a) We can see an upward trend over time but the slope of the trend differs during certain range periods. We can see that around 1991-1993, there are several dips in the workforce and the trend line is nearly parallel with the x-axis. Each month has different variation scales in the monthly seasonal graph.

b) The recession is not so apparent in the overall graph and the trend line, but it is noticeable in the remainder plot.

Problem 10)

This exercise uses the canadian_gas data (monthly Canadian gas production in billions of cubic metres, January 1960 – February 2005).

Plot the data using autoplot(), gg_subseries() and gg_season() to look at the effect of the changing seasonality over time.1 Do an STL decomposition of the data. You will need to choose a seasonal window to allow for the changing shape of the seasonal component. How does the seasonal shape change over time? [Hint: Try plotting the seasonal component using gg_season().] Can you produce a plausible seasonally adjusted series? Compare the results with those obtained using SEATS and X-11. How are they different?

a)

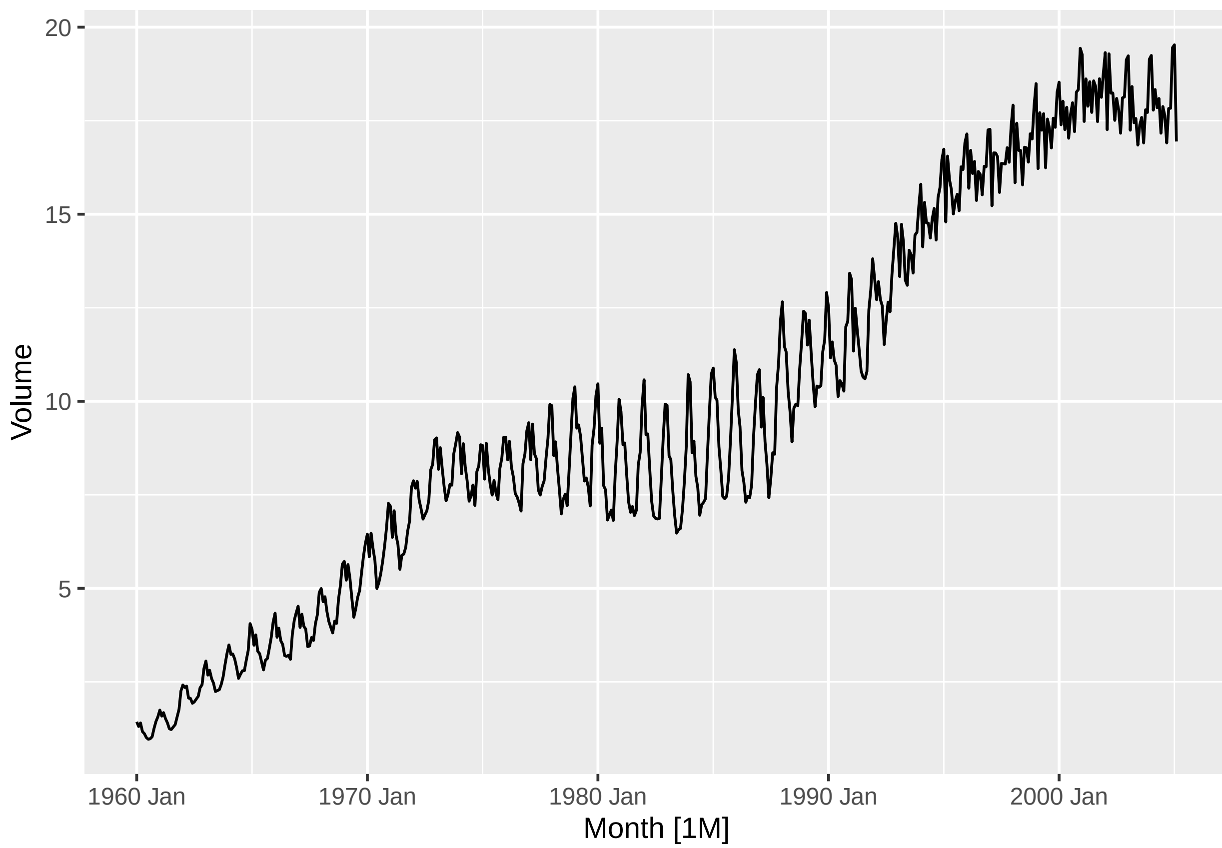
Code

```
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
```

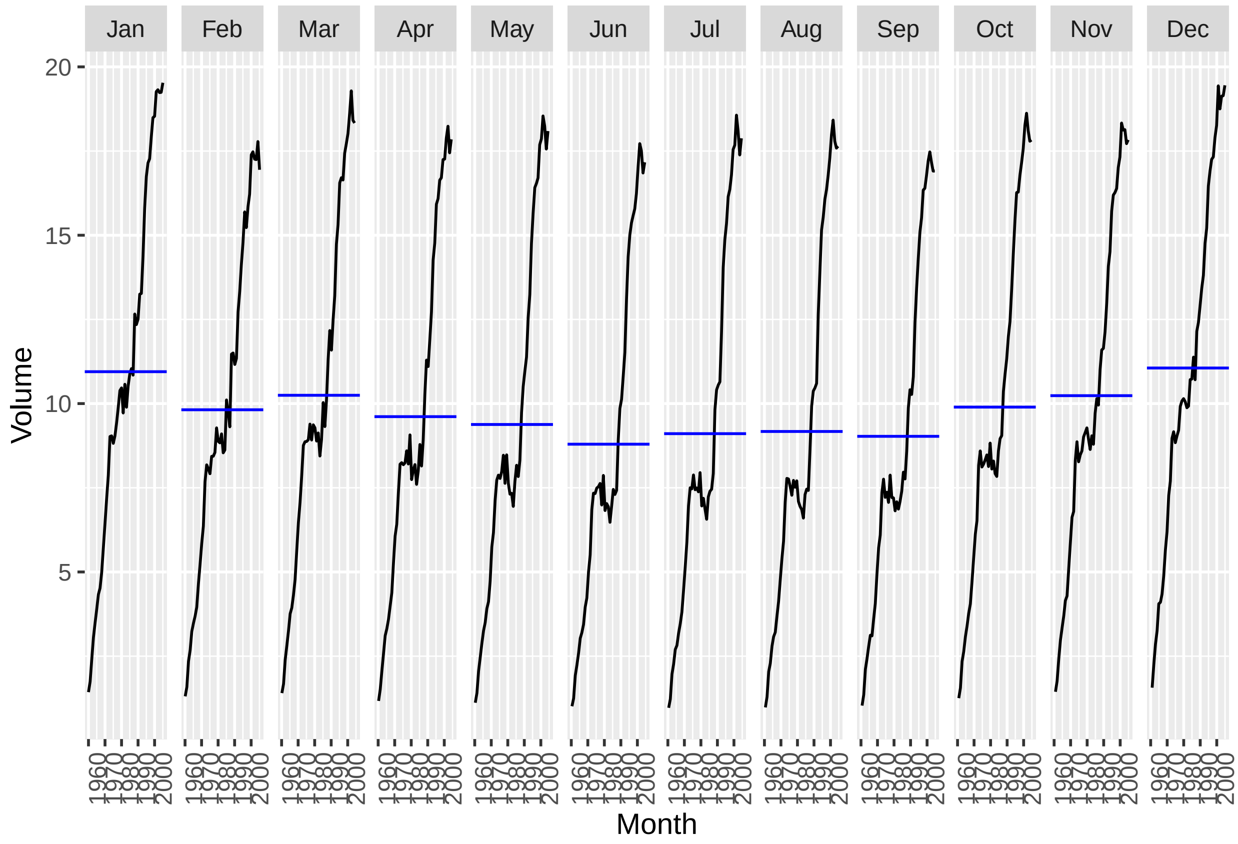
```
autoplot(canadian_gas)
```

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



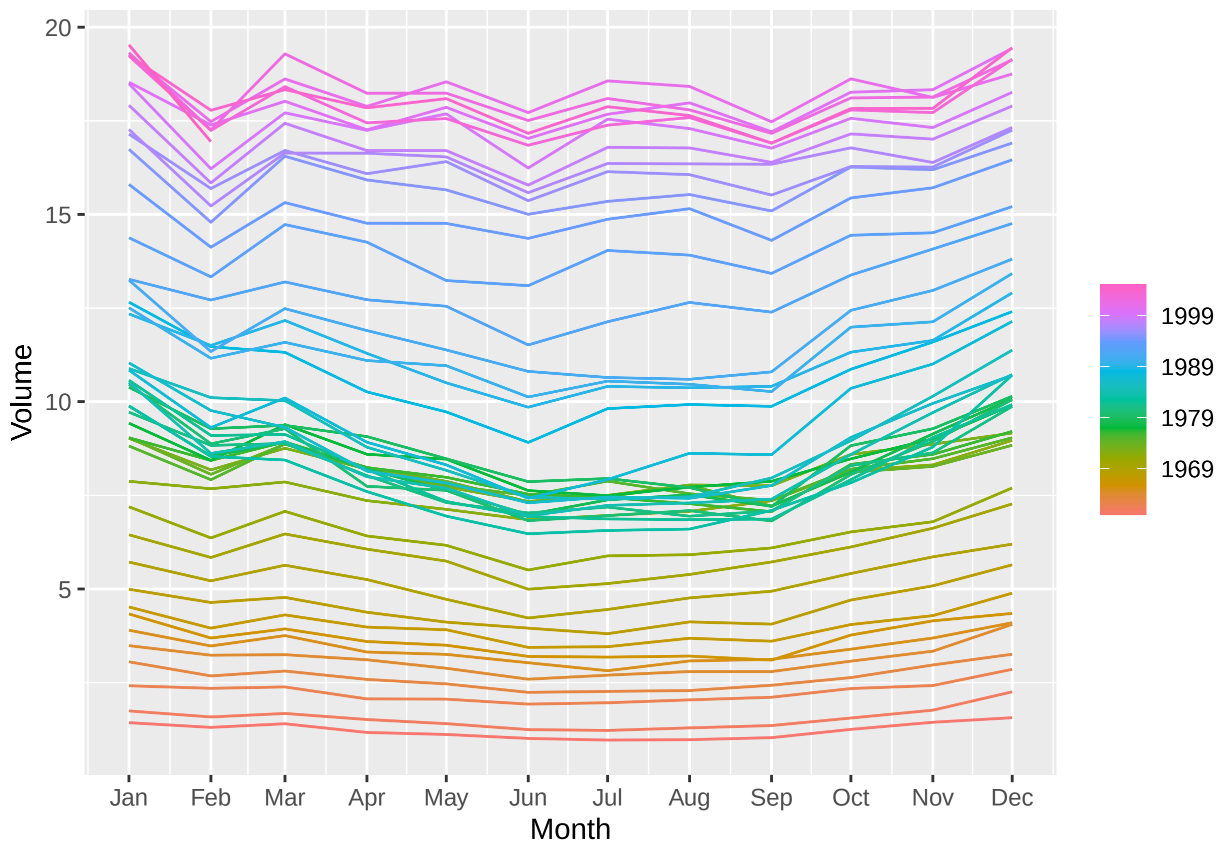

```
gg_subseries(canadian_gas)
```

```
## Plot variable not specified, automatically selected ‘y = Volume‘
```




```
gg_season(canadian_gas)
```

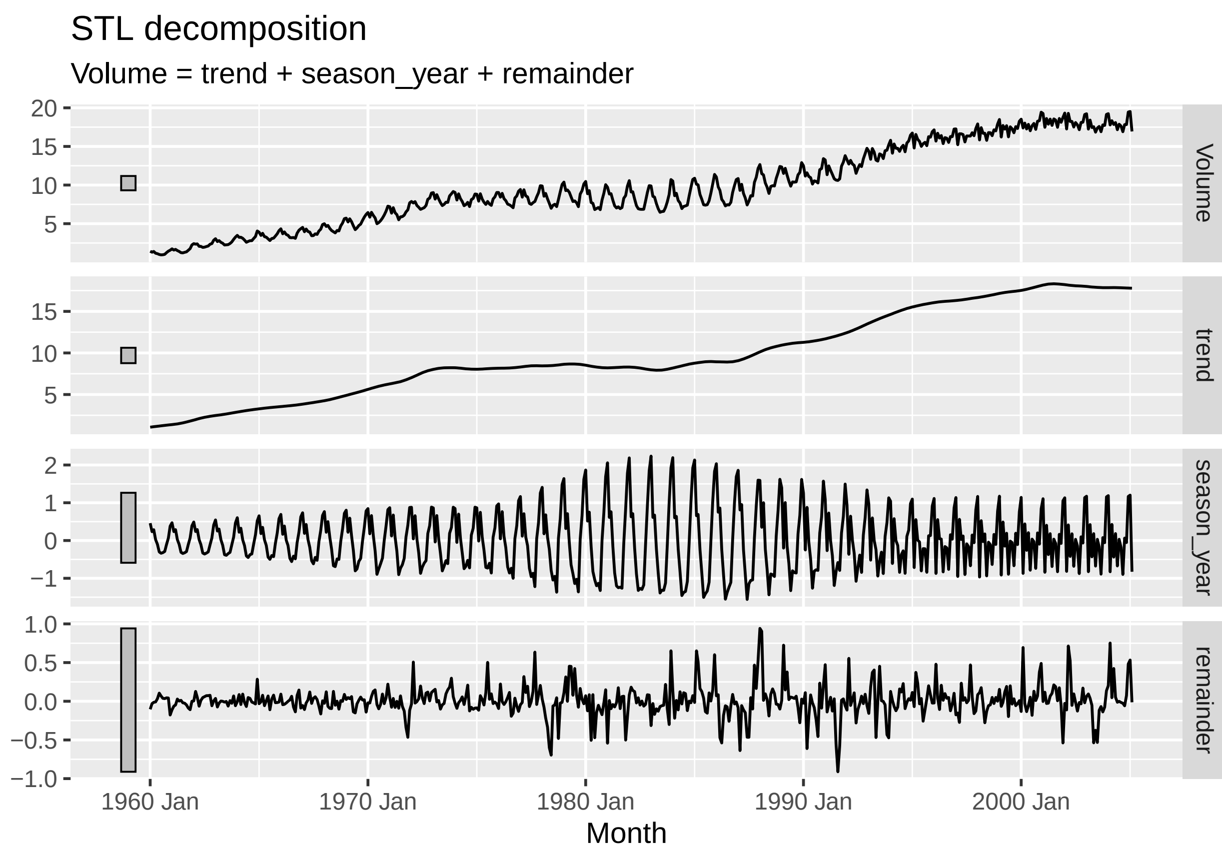
```
## Plot variable not specified, automatically selected `y = Volume`
```



b)

Code)

```
canadian_gas %>%
  model(STL(Volume ~ trend(window = 21) + season(window = 7), robust = TRUE)) %>%
  components() %>%
  autoplot()
```



c)

Answer)

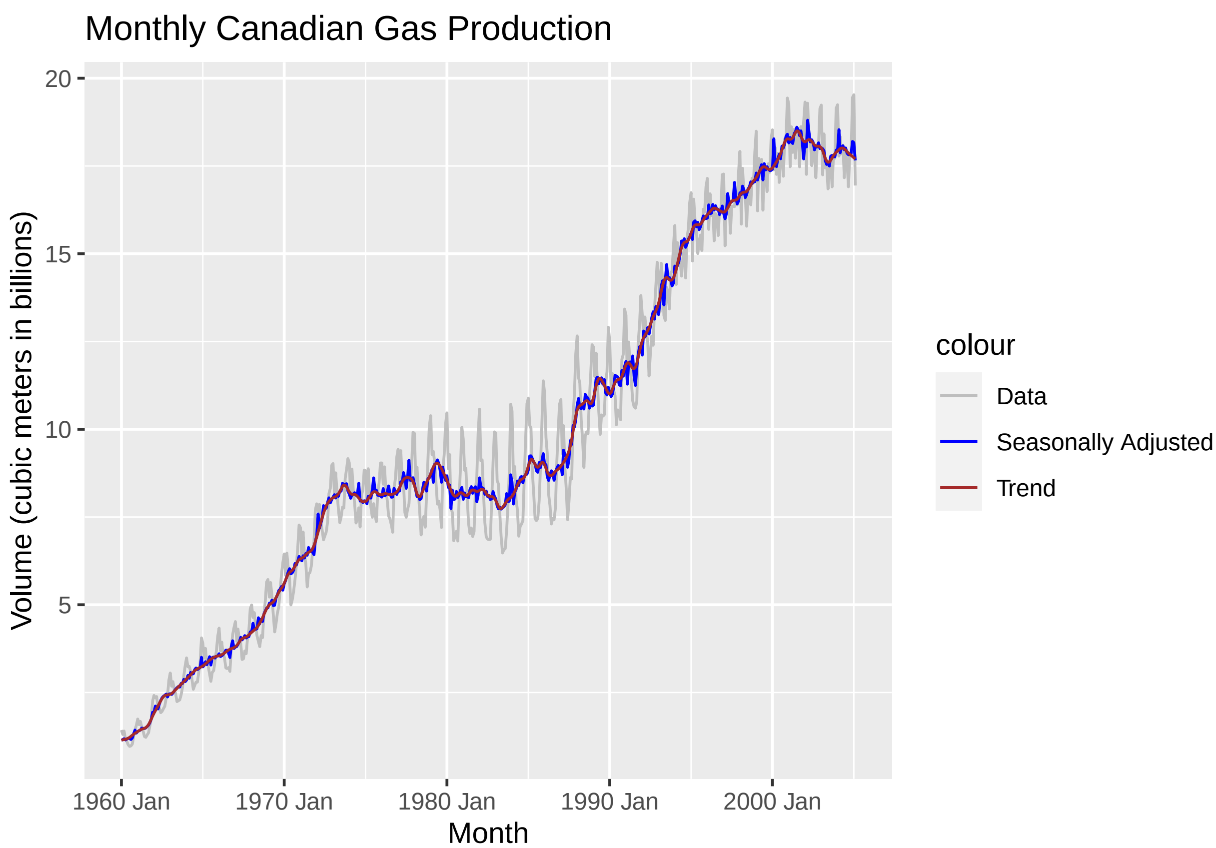
The seasonal shape seems to be identical with each other up until after 1975 and then the shape increases and then decreases back to almost normal during the mid 1990's. Around that time, there wasn't any trend with gas production.

d)

Code)

```
x11gas <- canadian_gas %>%
  model(x11 = X_13ARIMA_SEATS(Volume ~ x11())) %>%
  components()
```

```
x11gas %>%
  ggplot(aes(x = Month)) +
  geom_line(aes(y = Volume, colour = "Data")) +
  geom_line(aes(y = season_adjust, colour = "Seasonally Adjusted")) +
  geom_line(aes(y = trend, colour = "Trend")) +
  labs(y = "Volume (cubic meters in billions)", title = "Monthly Canadian Gas Production") +
  scale_colour_manual(values = c("gray", "blue", "brown"), breaks = c("Data", "Seasonally Adjusted", "T
```



e)

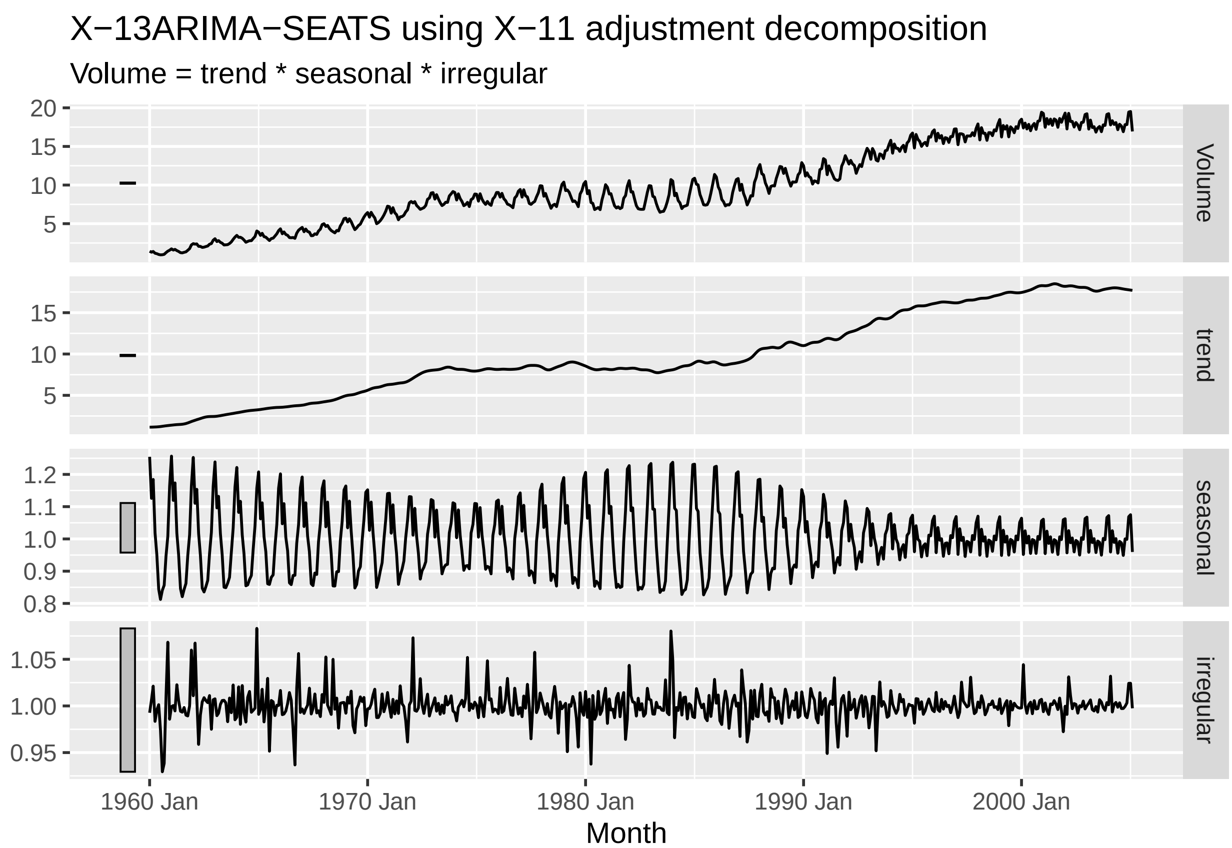
Answer)

The results from both X11 AND SEATS are pretty similar. The main difference is that the STL seasonal shape is unique from X11 and SEATS. Also, the remainder on the STL decomposition is the smallest compared to them all.

Code)

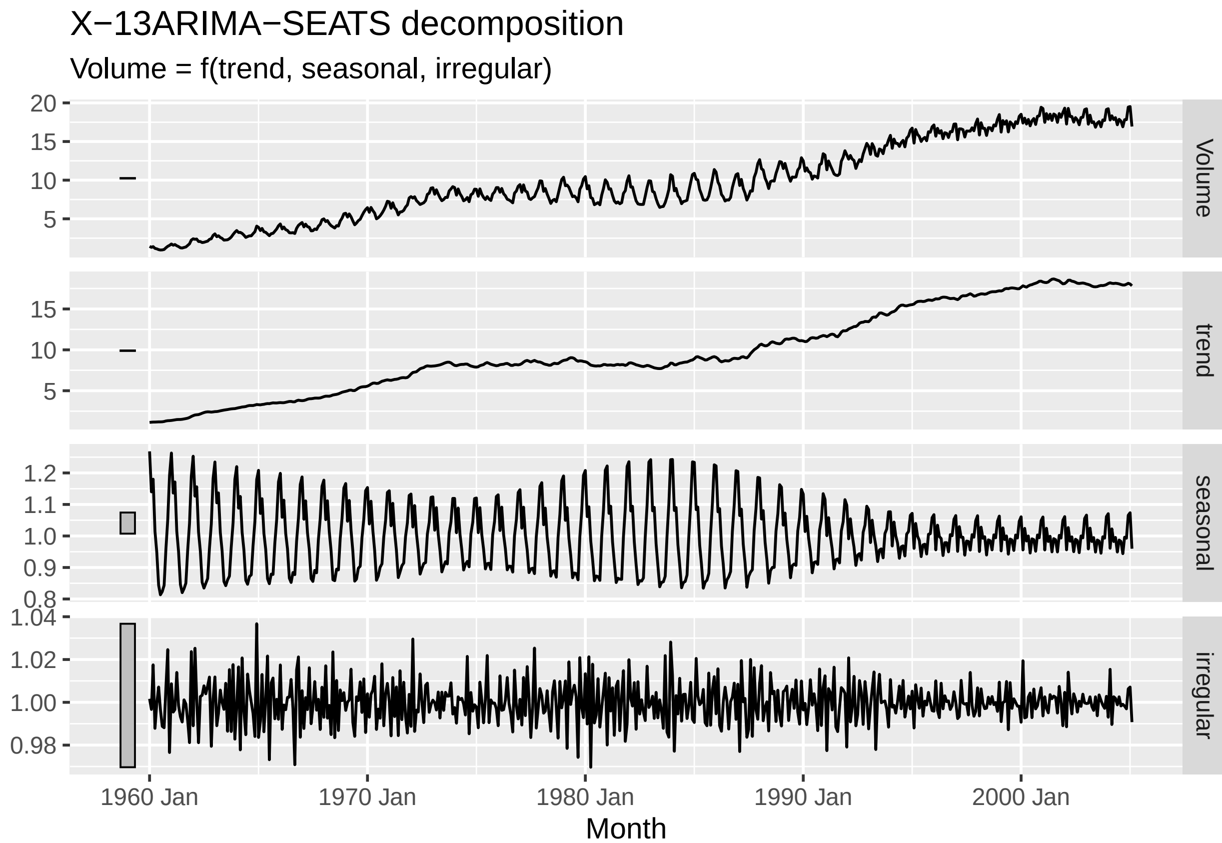
```
x11gas <- canadian_gas %>%
  model(x11 = X_13ARIMA_SEATS(Volume ~ x11())) %>%
  components()

autoplot(x11gas)
```




```r
seatsgas <- canadian_gas %>%
  model(seats = X_13ARIMA_SEATS(Volume ~ seats())) %>%
  components()

autoplot(seatsgas)
```



X−13ARIMA−SEATS decomposition
Volume = f(trend, seasonal, irregular)
20
15
10
5
Volume
15
10
5
trend
1.2
1.1
1.0
0.9
0.8
seasonal
1.04
1.02
1.00
0.98
irregular
1960 Jan
1970 Jan
1980 Jan
1990 Jan
2000 Jan
Month