621 hw3

2023-01-30

Problem 2

```r
library(fpp3)
```

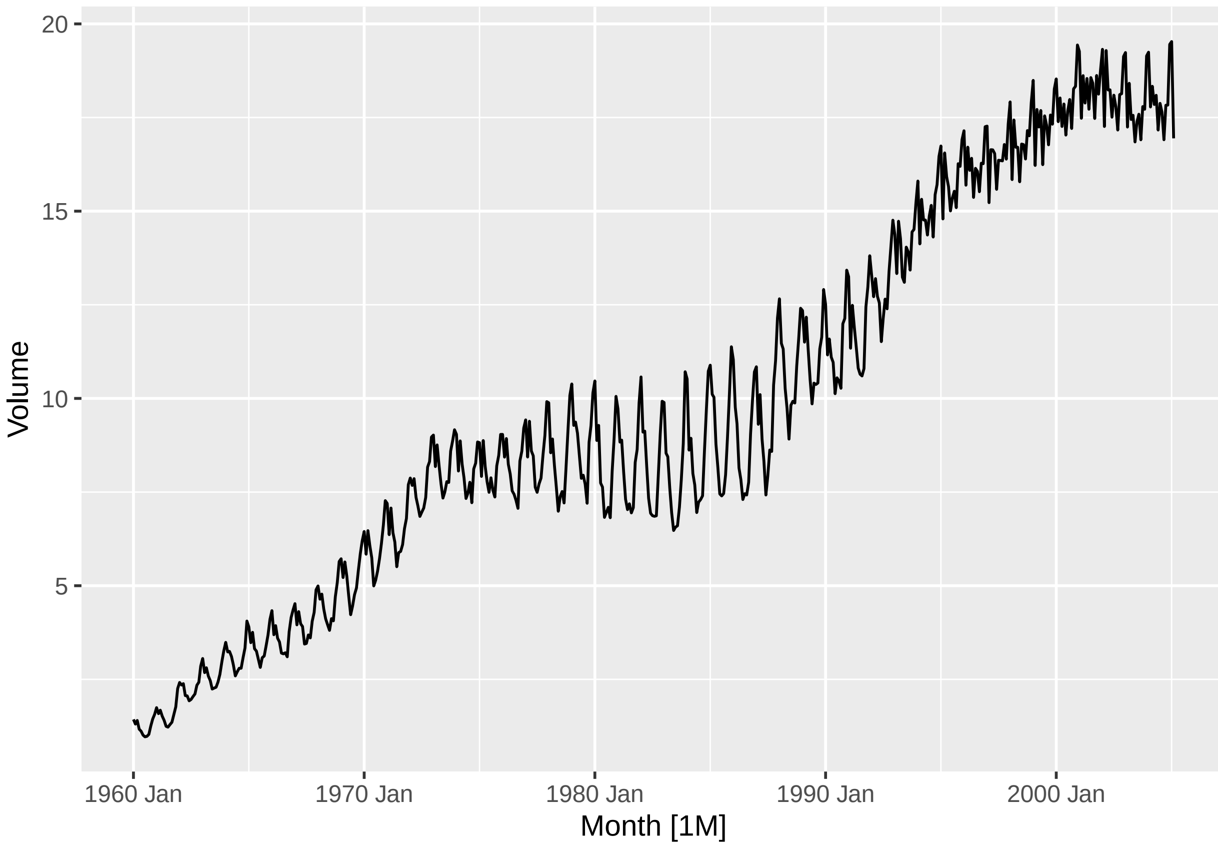
```
## -- Attaching packages -------------------------------------------- fpp3 0.4.0 --
## v tibble      3.1.8     v tsibble     1.1.3
## v dplyr       1.0.9     v tsibbledata 0.4.1
## v tidyr       1.2.0     v feasts      0.3.0
## v lubridate   1.8.0     v fable       0.3.2
## v ggplot2     3.3.6
## -- 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
```

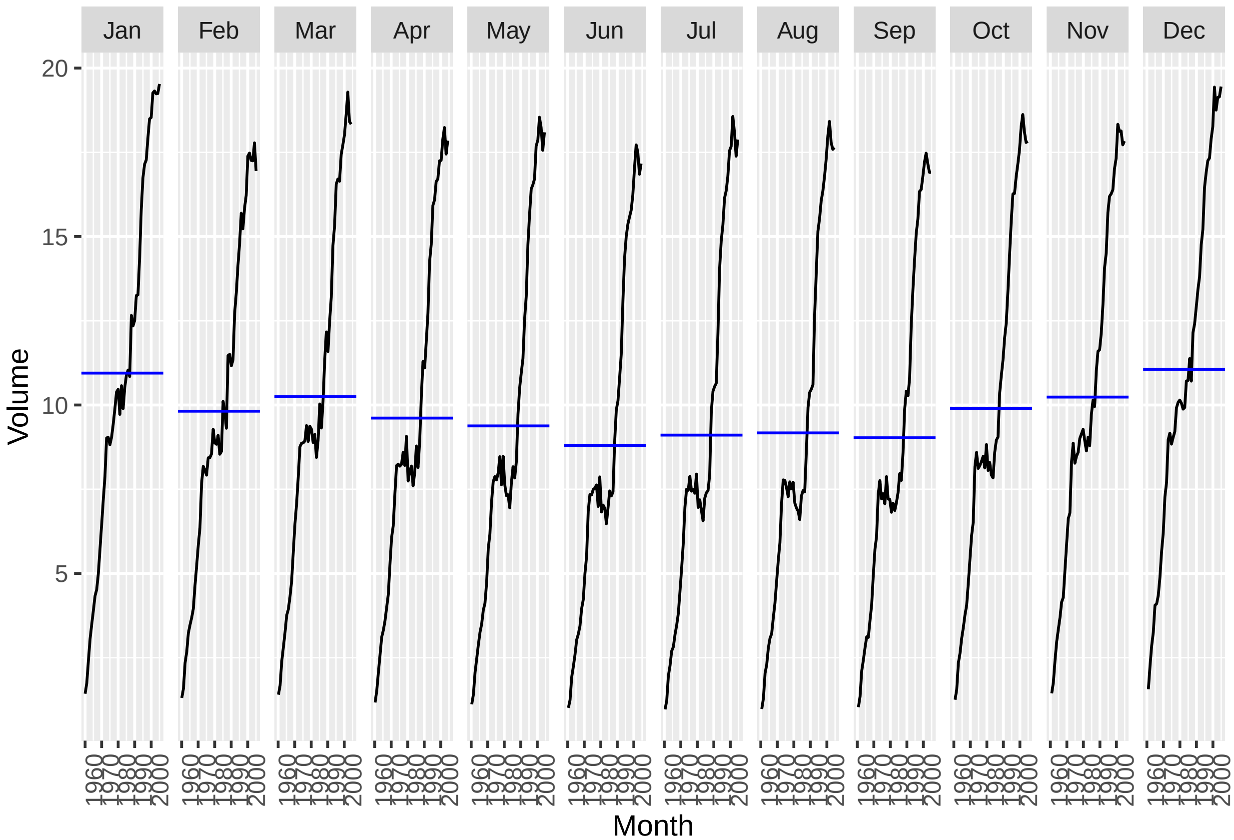
```r
library(ggfortify)
data(canadian_gas)
autoplot(canadian_gas)
```

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



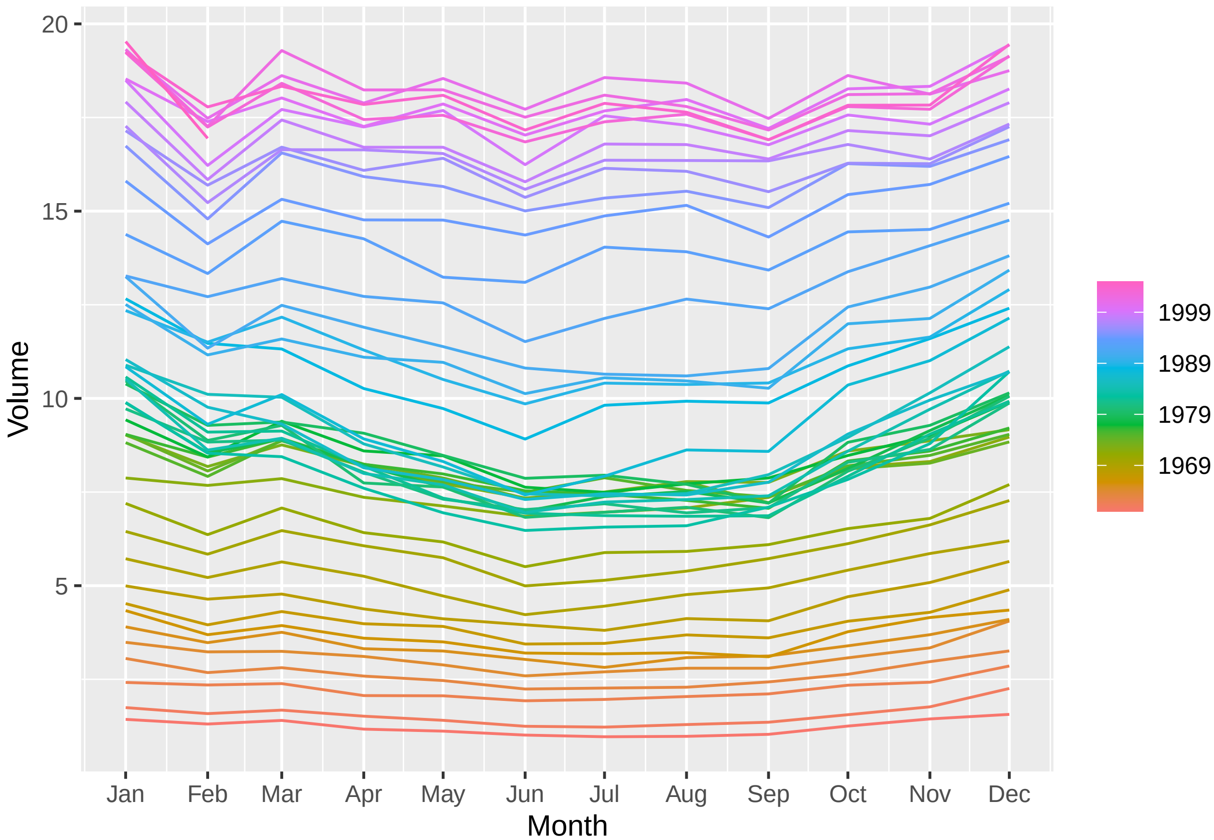

```
gg_subseries(canadian_gas)
```

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

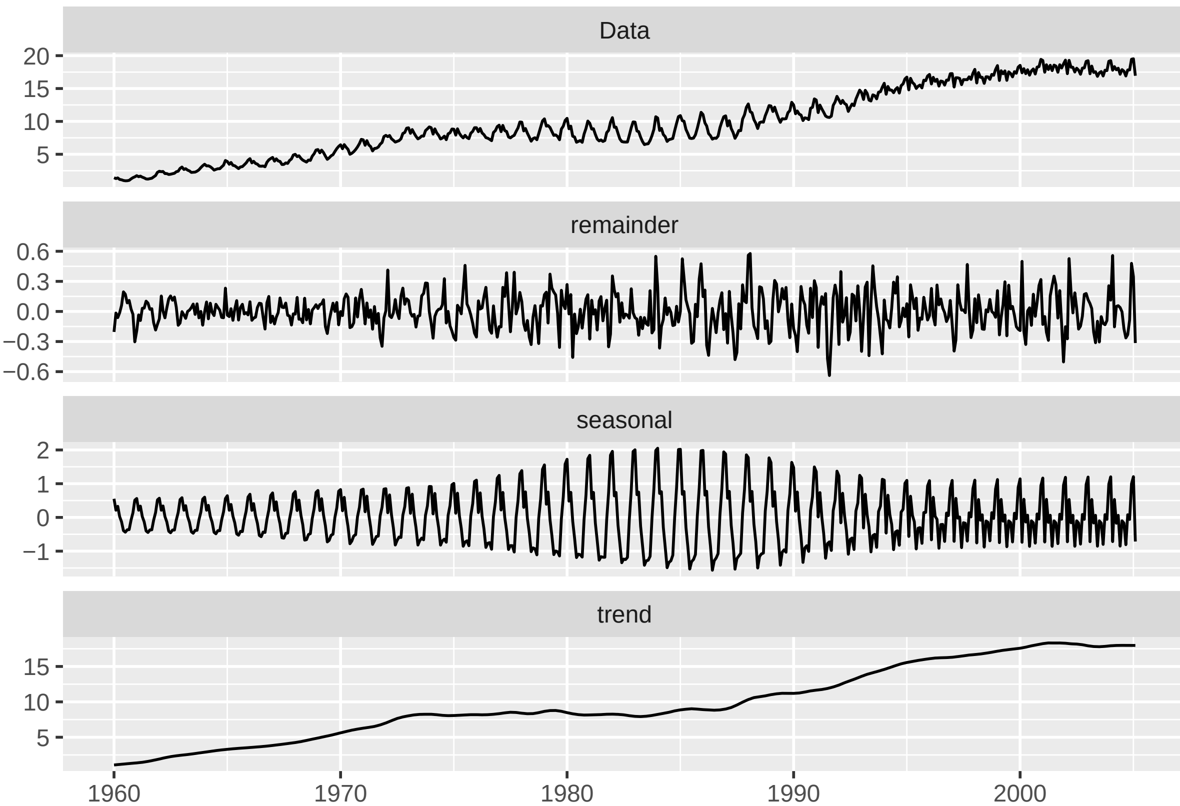



```
gg_season(canadian_gas)
```

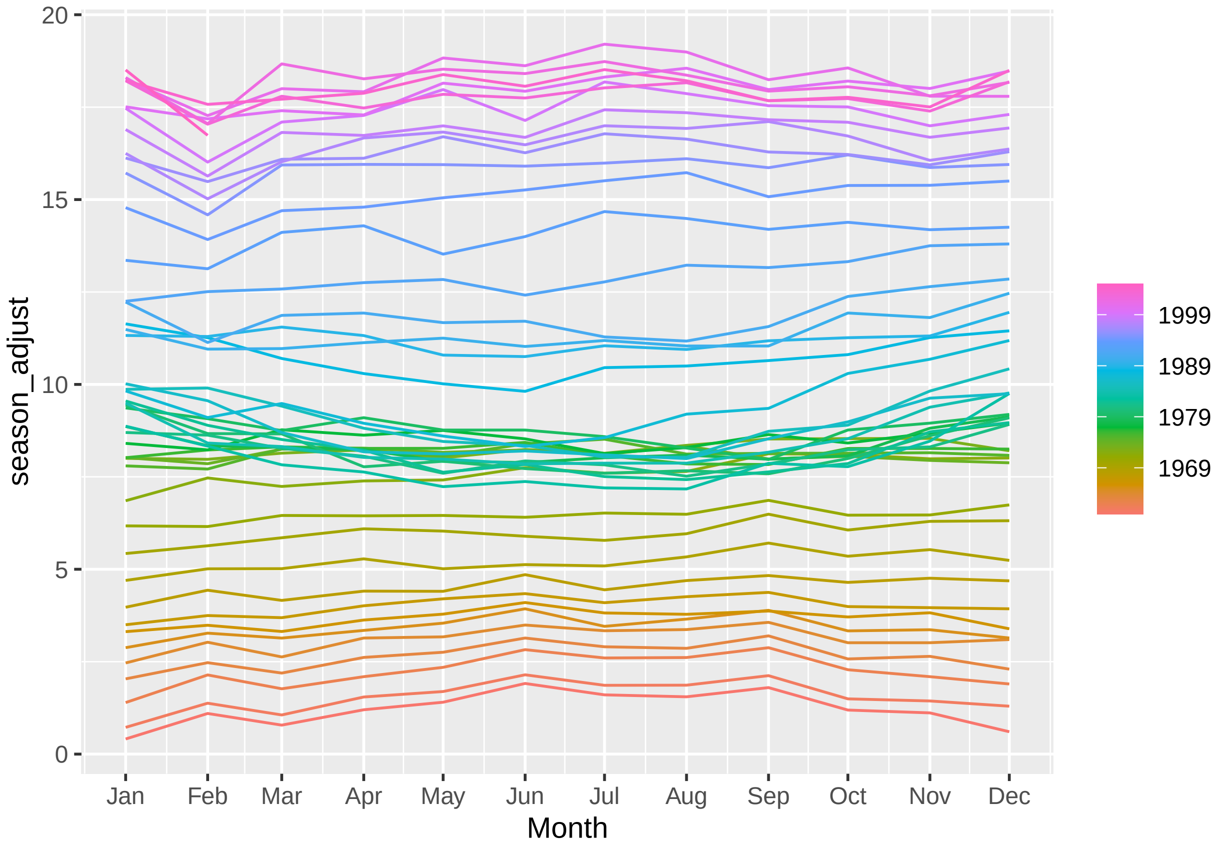
```
## Plot variable not specified, automatically selected `y = Volume`
```



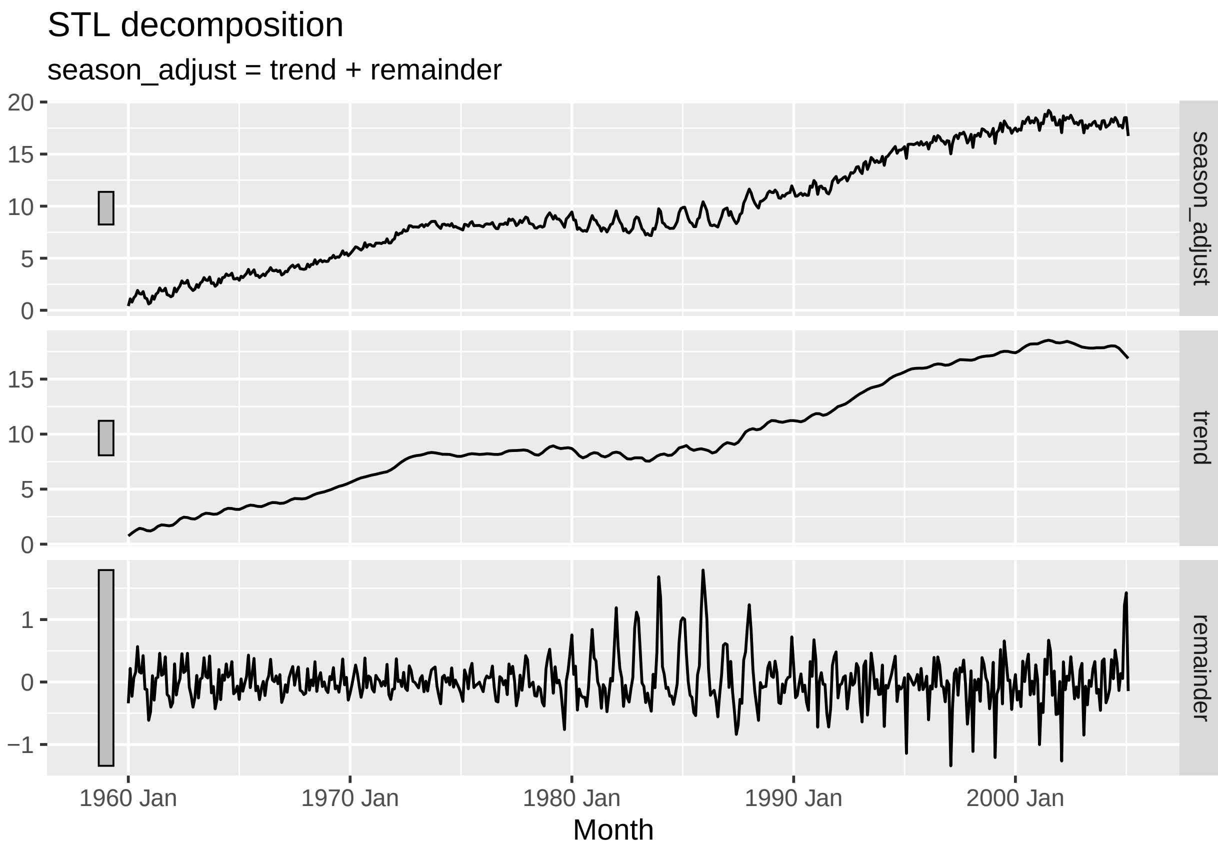

```
gas = stl(canadian_gas, s.window = 12)
autoplot(gas)
```



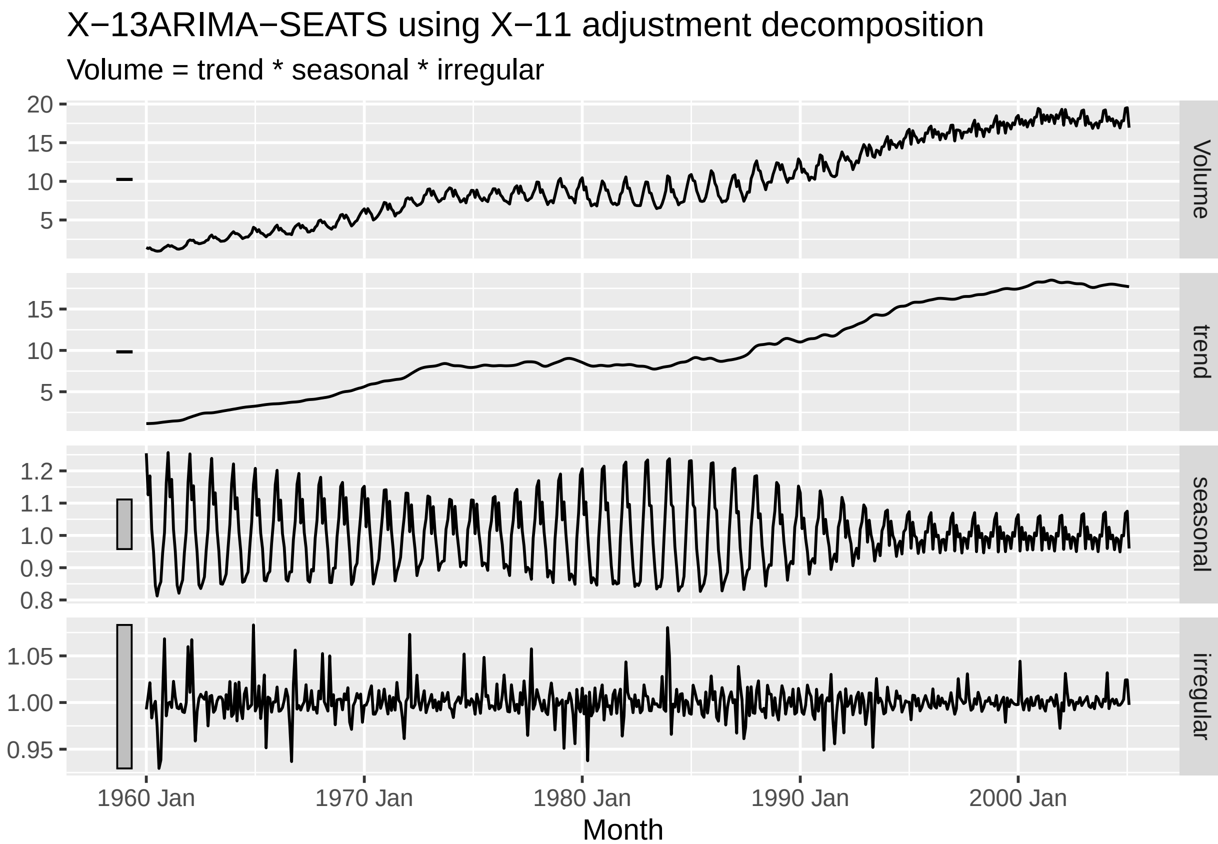

```
STL_gas <- canadian_gas %>%
  model(
    STL(Volume ~ trend(window=12) +
                 season(window="periodic"),
        robust = T)) %>%
  components()
STL_gas %>% gg_season(season_adjust)
```



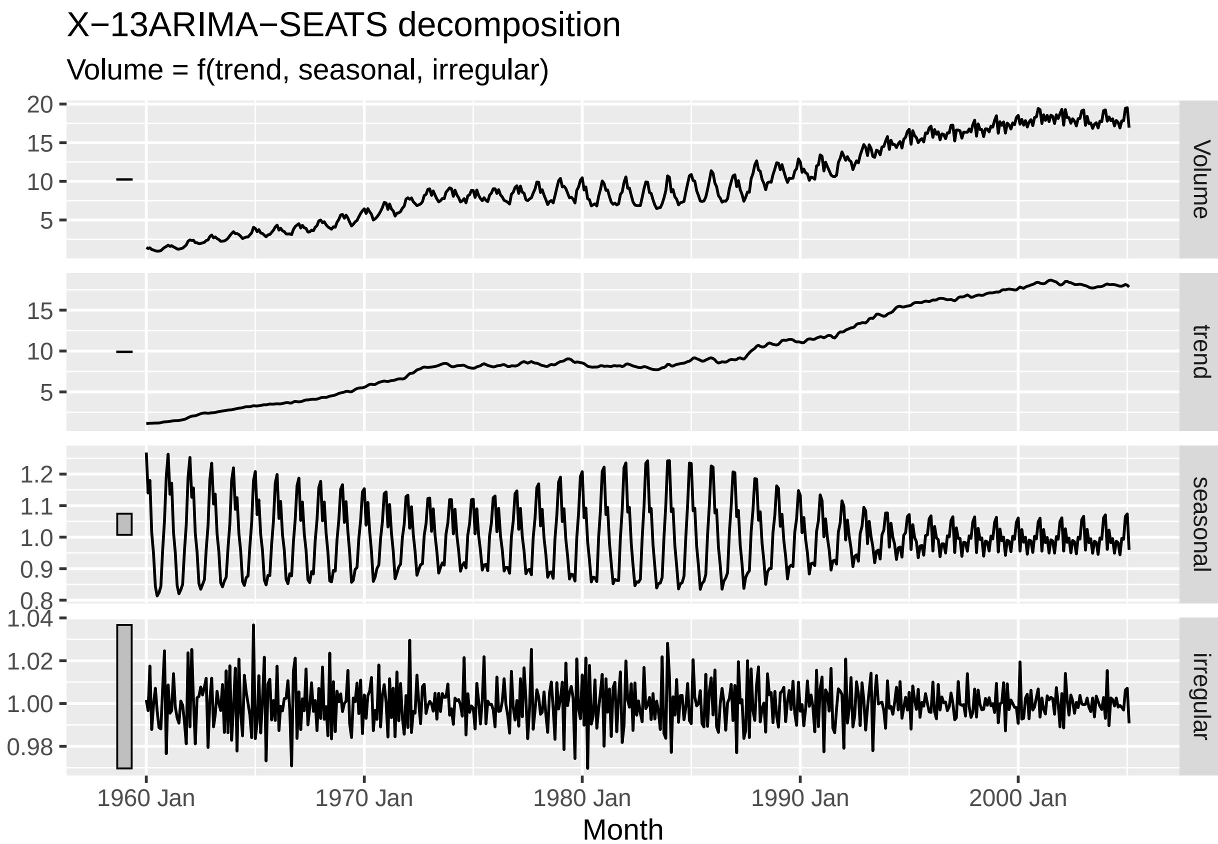

```
STL_gas %>%
  autoplot(season_adjust)
```




```
x11_gas <- canadian_gas %>%
  model( x11 = X_13ARIMA_SEATS(Volume ~ x11())) %>%
  components()
autoplot(x11_gas)
```




```
seats_gas <- canadian_gas %>%
  model(seats = X_13ARIMA_SEATS(Volume ~ seats())) %>%
  components()
autoplot(seats_gas)
```



According to the plot, there is an increasing trend between 1960 and 2000. Besides, there is an obvious increase during winter. After STL decomposition, the variability becomes larger. x11 and SEATS have different seasonal variance.

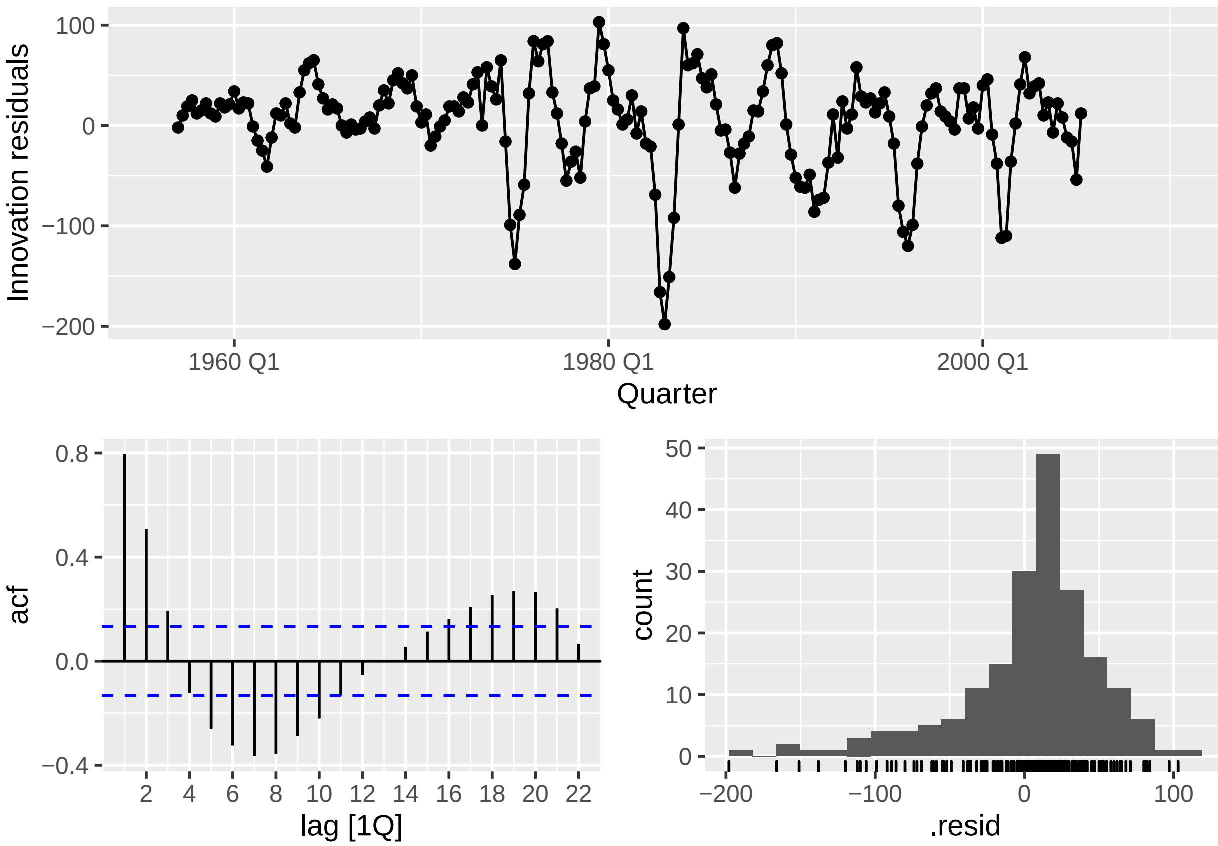
Problem 3

```
bricks_production <- aus_production %>%
  select(Bricks)
fit <- bricks_production %>% model(SNAIVE(Bricks))
fit %>% gg_tsresiduals()
```

```
## Warning: Removed 24 row(s) containing missing values (geom_path).
```

```
## Warning: Removed 24 rows containing missing values (geom_point).
```

```
## Warning: Removed 24 rows containing non-finite values (stat_bin).
```



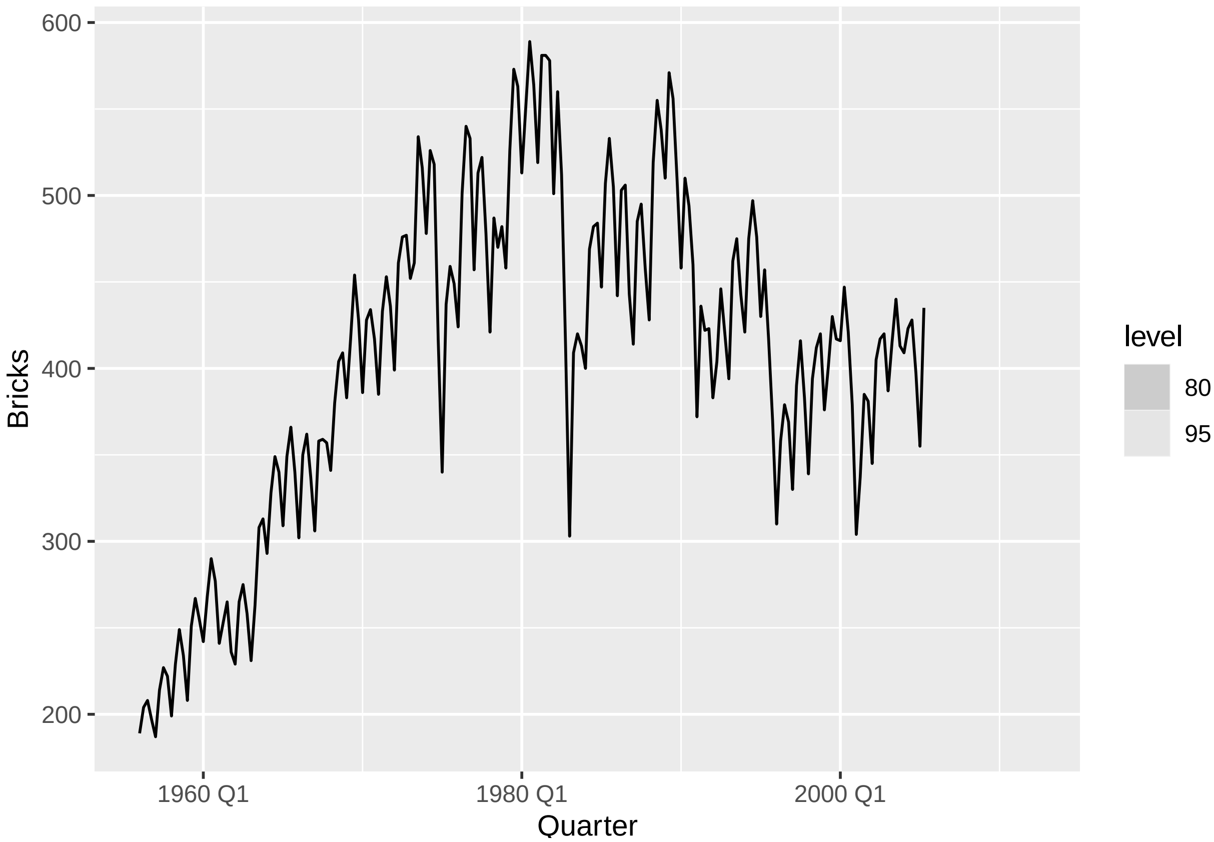

```
fit %>% forecast() %>% autoplot(bricks_production)
```

```
## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf

## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf

## Warning: Removed 8 row(s) containing missing values (geom_path).

## Warning: Removed 20 row(s) containing missing values (geom_path).
```



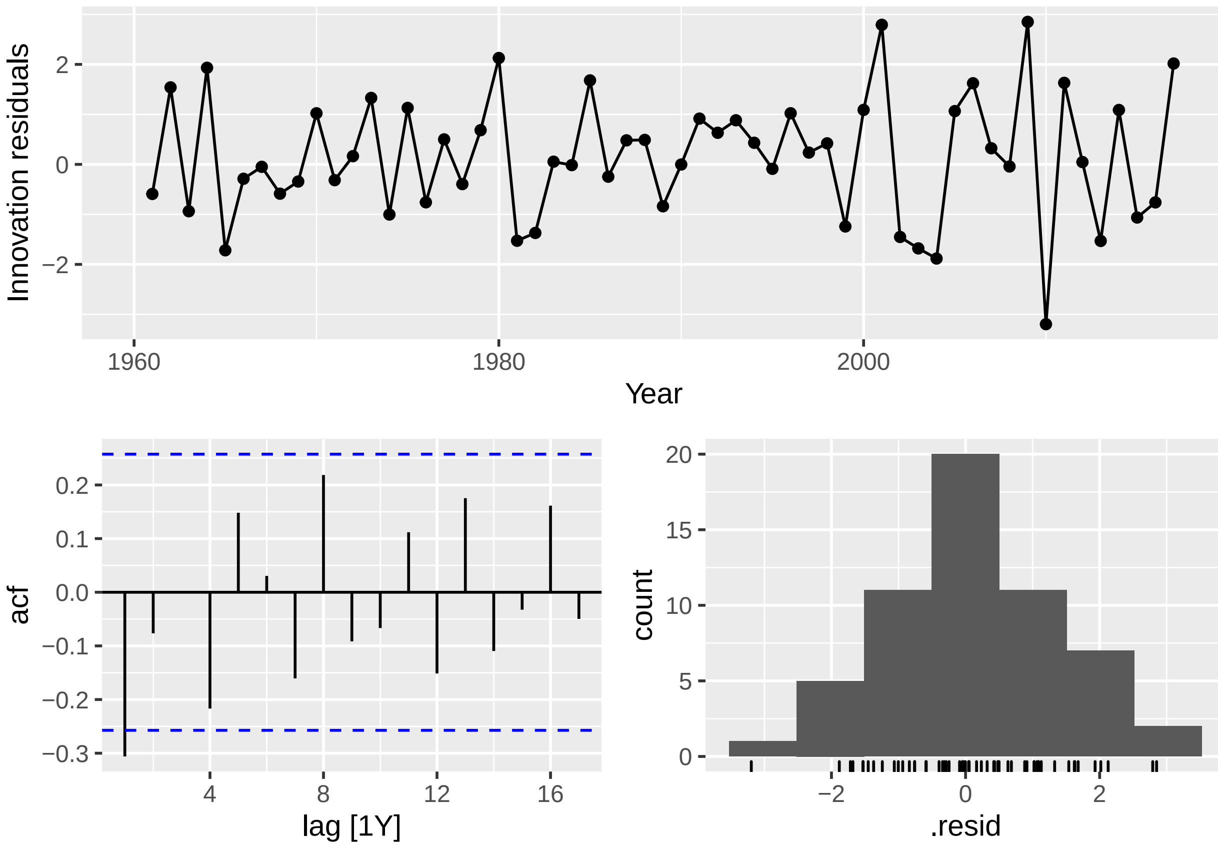

```
exports <- global_economy %>%
filter(Year>= 1960, Country == "Australia")

fit <- exports %>% model(NAIVE(Exports ))
fit %>% gg_tsresiduals()
```

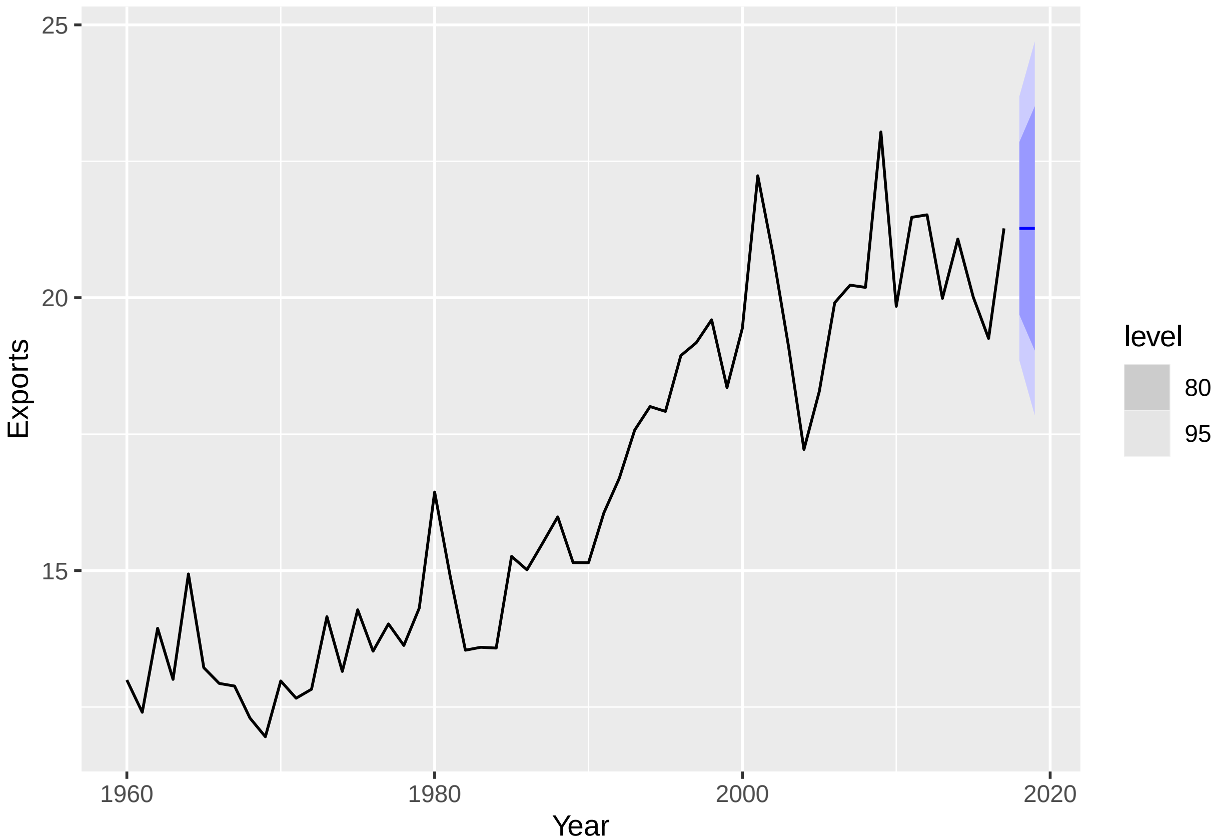
```
## Warning: Removed 1 row(s) containing missing values (geom_path).

## Warning: Removed 1 rows containing missing values (geom_point).

## Warning: Removed 1 rows containing non-finite values (stat_bin).
```

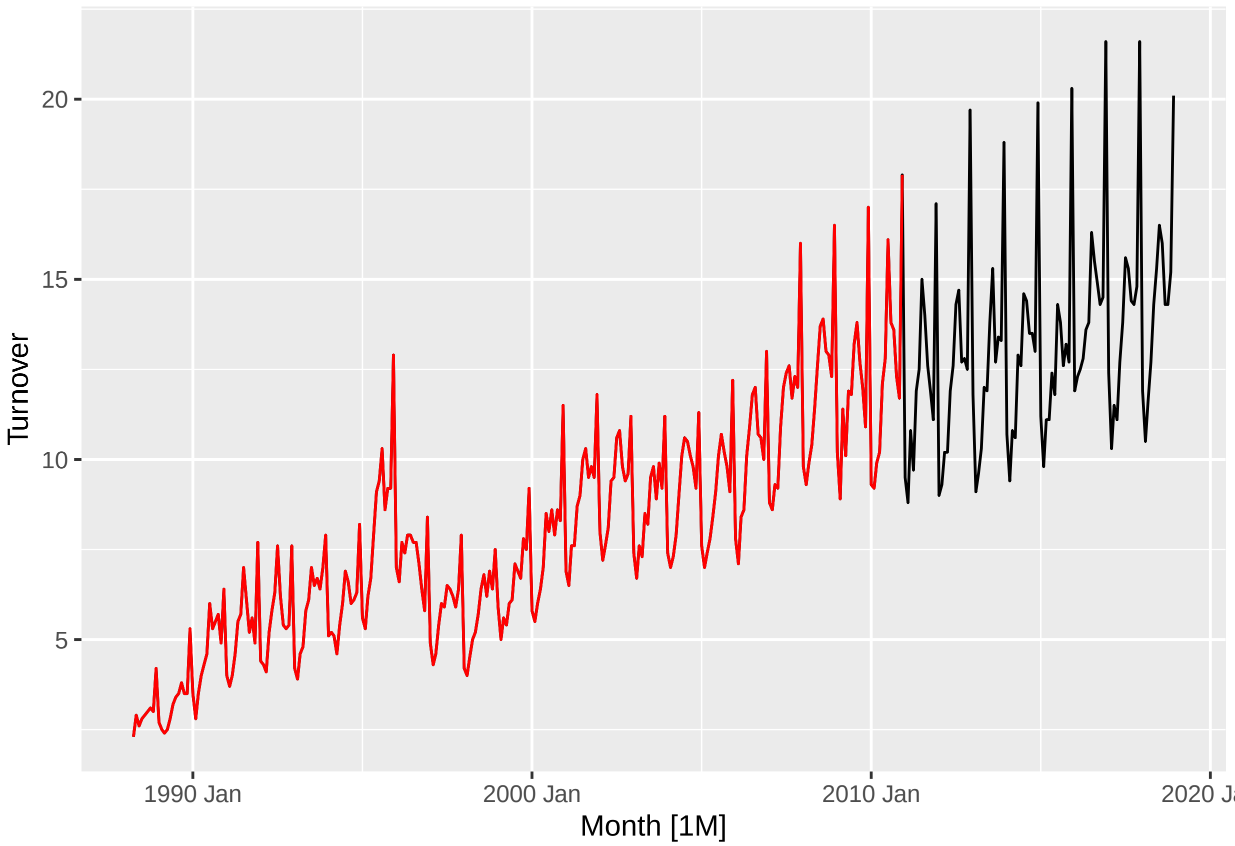



```
fit %>% forecast() %>% autoplot(exports)
```



Problem 5

```
set.seed(12345678)
myseries <- aus_retail %>%
  filter(`Series ID` == sample(aus_retail$`Series ID`,1))
myseries_train <- myseries %>%
  filter(year(Month) < 2011)
autoplot(myseries, Turnover) +
  autolayer(myseries_train, Turnover, colour = "red")
```

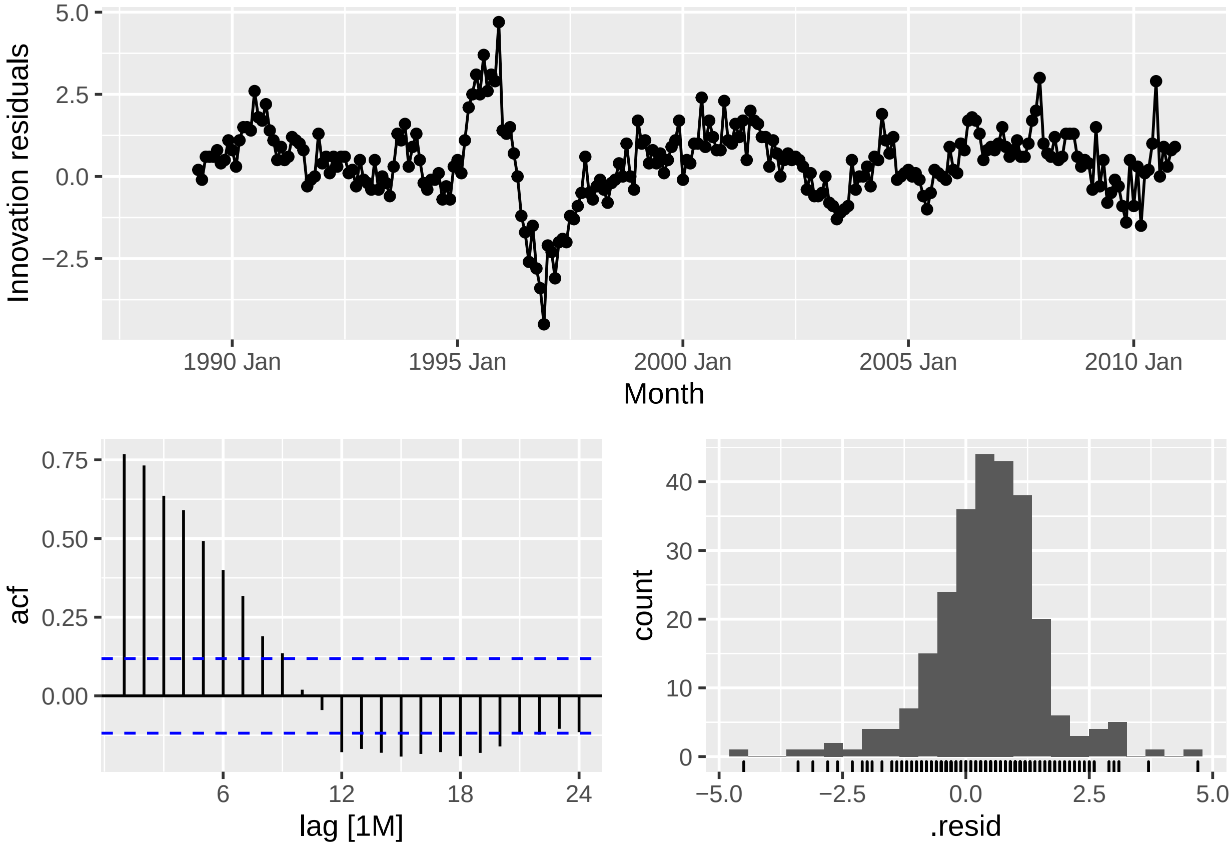



```
fit <- myseries_train %>%
          model(SNAIVE(Turnover))
fit %>% gg_tsresiduals()
```

```
## Warning: Removed 12 row(s) containing missing values (geom_path).
```

```
## Warning: Removed 12 rows containing missing values (geom_point).
```

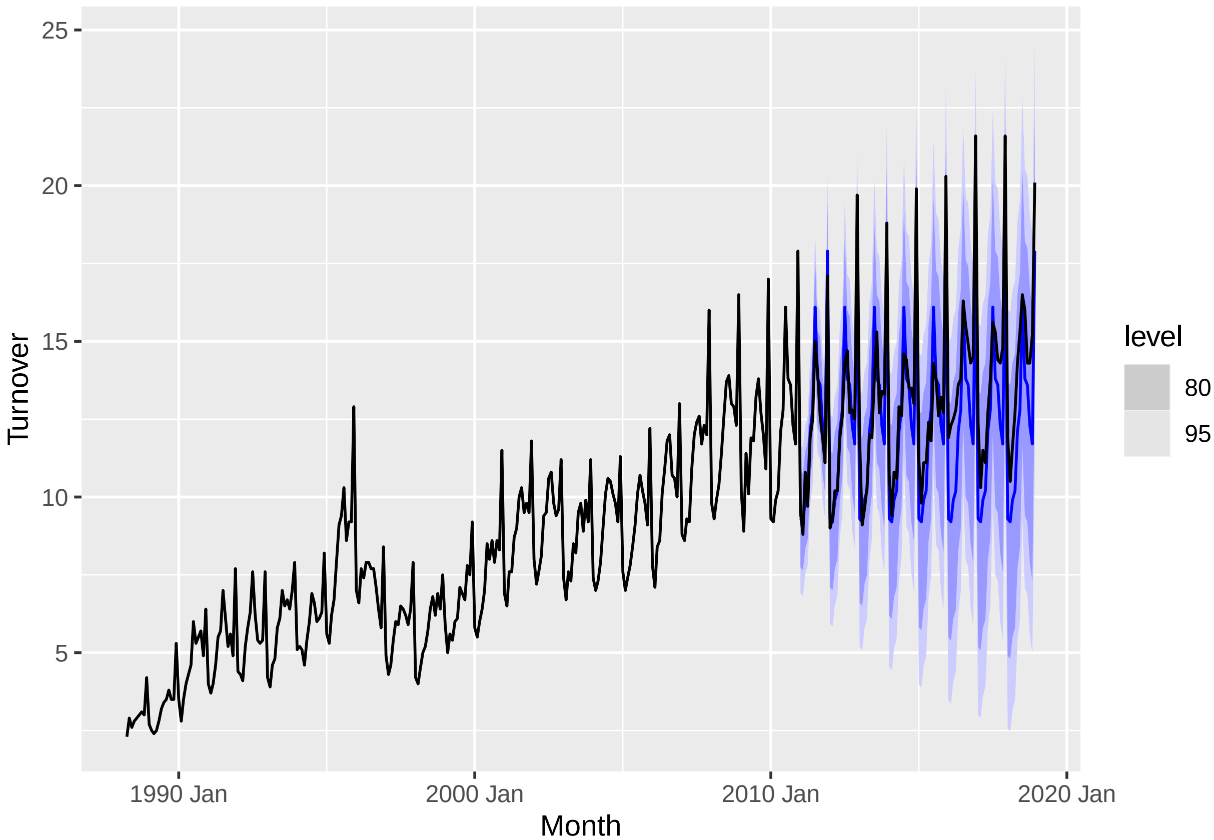
```
## Warning: Removed 12 rows containing non-finite values (stat_bin).
```




```
fc <- fit %>%
  forecast(new_data = anti_join(myseries, myseries_train))
```

```
## Joining, by = c("State", "Industry", "Series ID", "Month", "Turnover")
```

```
fc %>% autoplot(myseries)
```




```
fit %>% accuracy()
```

```
## # A tibble: 1 x 12
##   State     Indus~1 .model .type    ME  RMSE   MAE   MPE  MAPE  MASE RMSSE  ACF1
##   <chr>     <chr>   <chr>  <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 Northern~ Clothi~ SNAIV~ Trai~ 0.439  1.21 0.915  5.23  12.4     1     1 0.768
## # ... with abbreviated variable name 1: Industry
```

```
fc %>% accuracy(myseries)
```

```
## # A tibble: 1 x 12
##   .model      State Indus~1 .type    ME  RMSE   MAE   MPE  MAPE  MASE RMSSE  ACF1
##   <chr>       <chr> <chr>   <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 SNAIVE(Tu~ Nort~ Clothi~ Test  0.836  1.55  1.24  5.94  9.06  1.36  1.28 0.601
## # ... with abbreviated variable name 1: Industry
```

From 1995 to 2000, the residuals are worse. According to the residual plot, it has a shape of normal distribution, but it does not centered in 0.

It is important to choose appropriate data consisting of training set. If there are many similar data, the accuracy of forcesting is poor, if there are less various data, there may be some important information we will ignore.

Problem 6

```
library(astsa)
```

```
##
```

```
## Attaching package: 'astsa'

## The following object is masked _by_ '.GlobalEnv':
##
##     gas

## The following objects are masked from 'package:seasonal':
##
##     trend, unemp
```

```r
trend = time(jj) - 1970
Q = factor(cycle(jj) )
reg1 = lm(log(jj)~0 + trend + Q, na.action=NULL)
summary(reg1)
```

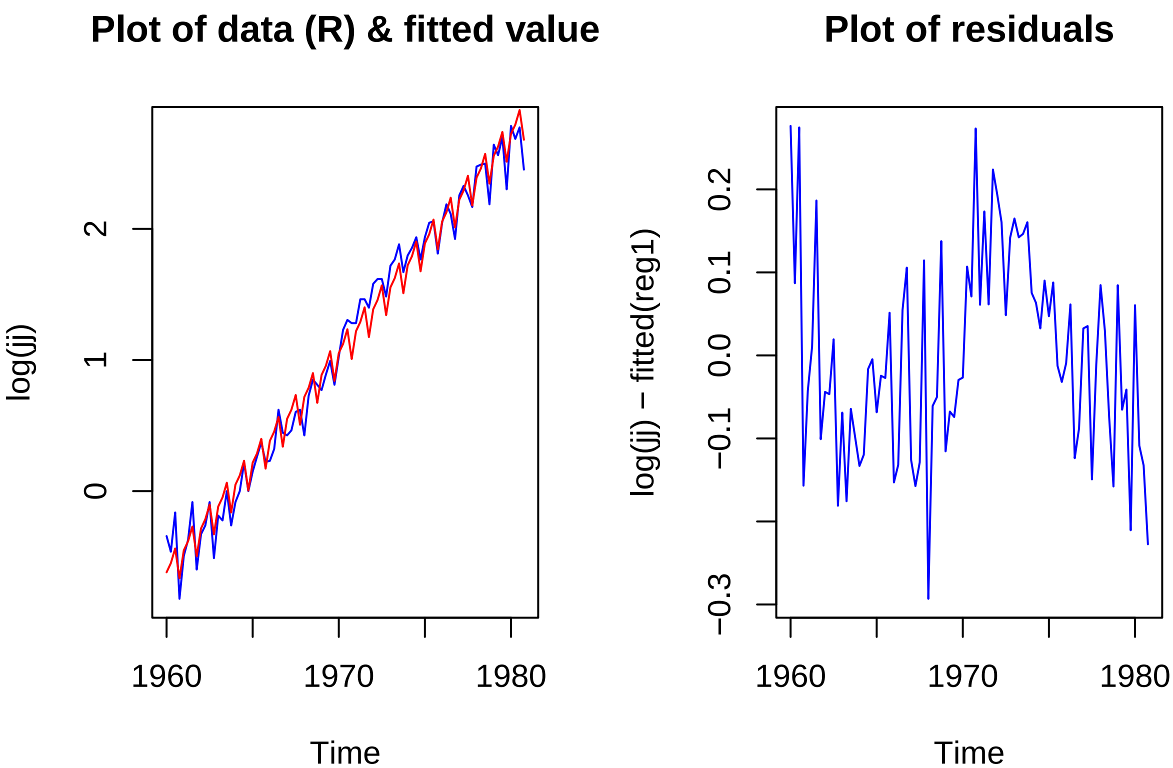
```
##
## Call:
## lm(formula = log(jj) ~ 0 + trend + Q, na.action = NULL)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.29318 -0.09062 -0.01180  0.08460  0.27644
##
## Coefficients:
##       Estimate Std. Error t value Pr(>|t|)
## trend 0.167172   0.002259   74.00   <2e-16 ***
## Q1    1.052793   0.027359   38.48   <2e-16 ***
## Q2    1.080916   0.027365   39.50   <2e-16 ***
## Q3    1.151024   0.027383   42.03   <2e-16 ***
## Q4    0.882266   0.027412   32.19   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1254 on 79 degrees of freedom
## Multiple R-squared:  0.9935, Adjusted R-squared:  0.9931
## F-statistic:  2407 on 5 and 79 DF,  p-value: < 2.2e-16
```

```r
reg2 = lm(log(jj)~trend + Q, na.action=NULL)
summary(reg2)
```

```
##
## Call:
## lm(formula = log(jj) ~ trend + Q, na.action = NULL)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.29318 -0.09062 -0.01180  0.08460  0.27644
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.052793   0.027359  38.480  < 2e-16 ***
## trend        0.167172   0.002259  73.999  < 2e-16 ***
## Q2           0.028123   0.038696   0.727   0.4695
## Q3           0.098231   0.038708   2.538   0.0131 *
## Q4          -0.170527   0.038729  -4.403 3.31e-05 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1254 on 79 degrees of freedom
## Multiple R-squared:  0.9859, Adjusted R-squared:  0.9852
## F-statistic:  1379 on 4 and 79 DF,  p-value: < 2.2e-16
```

```r
par(mfrow=c(1,2))
plot(log(jj), main="Plot of data (R) & fitted value", col="blue")
lines(fitted(reg1), col="red")
plot(log(jj) - fitted(reg1), main="Plot of residuals", col="blue")
```



If there is an intercept in the model, there will not exist Q1. According to the plot, the residuals are fluctuating around 0, which means it fitted well.