

ShovanBiswas-DATA24_Homework_02

Shovan Biswas

9/12/2020

Libraries

```
library(knitr)
library(kableExtra)
library(tidyverse)
library(gridExtra)
library(fpp2)
library(readxl)
```

Exercise 3.1

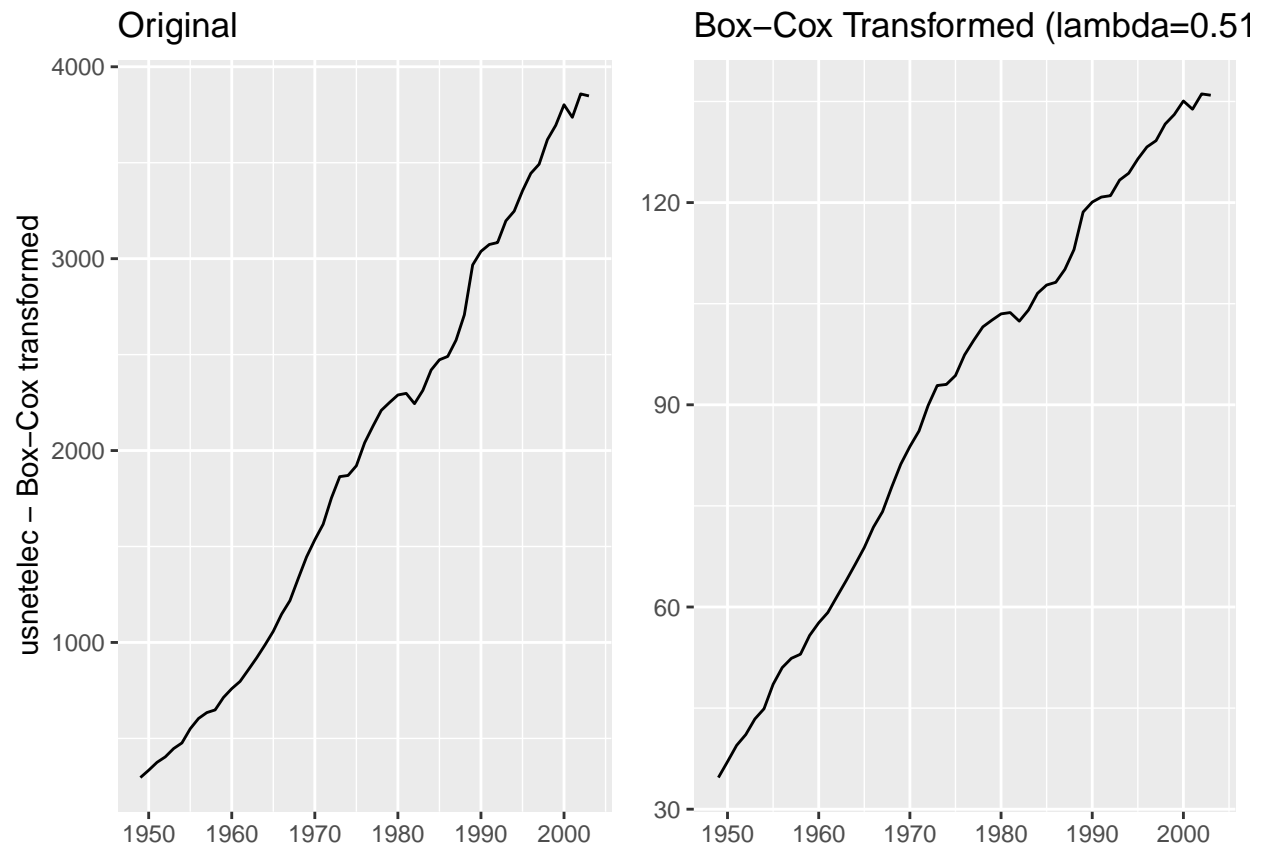
For the following series, find an appropriate Box-Cox transformation in order to stabilise the variance.

Since there are four items, I created a loop to contrast the original with the Box-Cox transformation of the data.

```
boxcox_trans_loop <- function(item, title){
  lambda <- BoxCox.lambda(item)
  Visualization1 <- autoplot(item) + ggtitle("Original") + ylab(title) + theme(axis.title.x = element_blank())
  Visualization2 <- autoplot(BoxCox(item, lambda)) + ggtitle(paste0("Box-Cox Transformed (lambda=", round(lambda, 2), ")"))
  grid.arrange(Visualization1, Visualization2, ncol = 2)
}
```

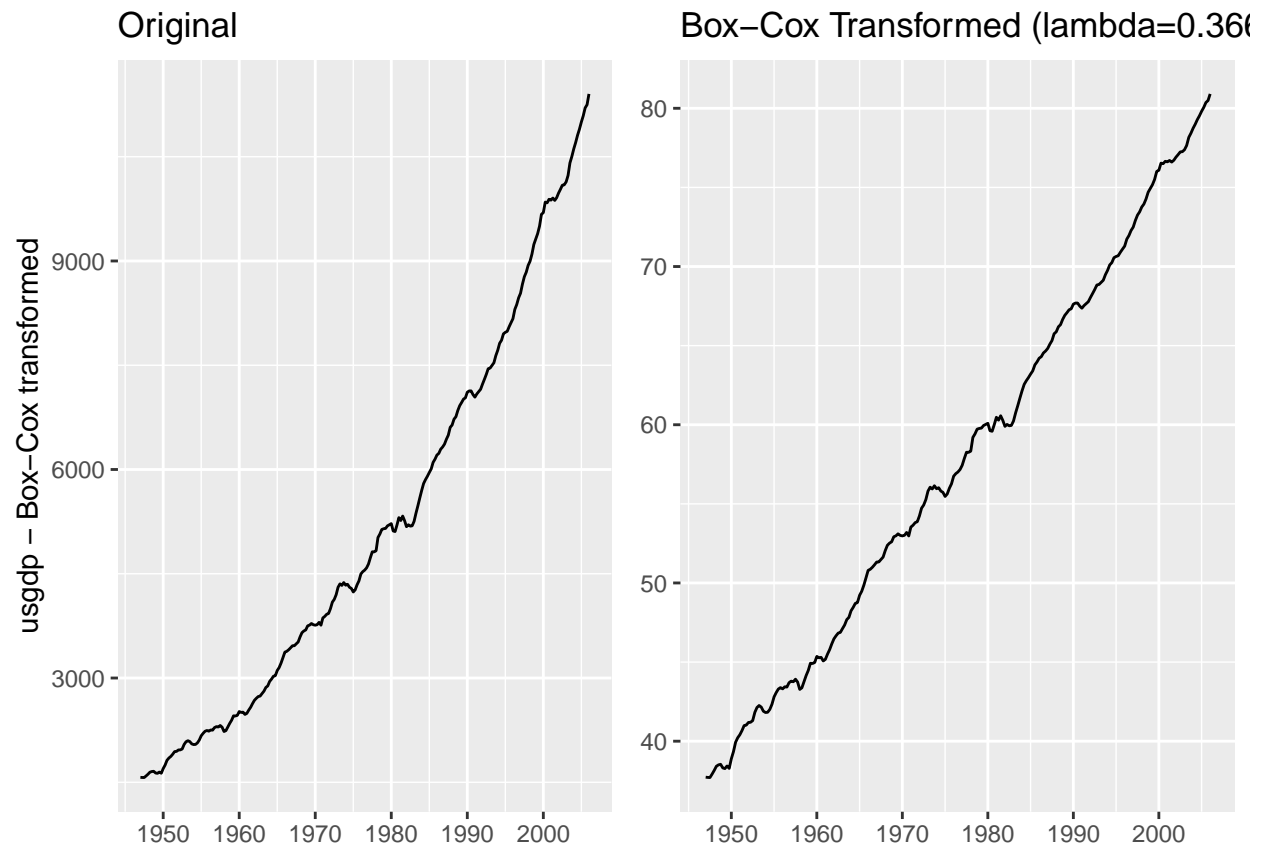
usnetelec

```
boxcox_trans_loop(usnetelec, "usnetelec - Box-Cox transformed")
```



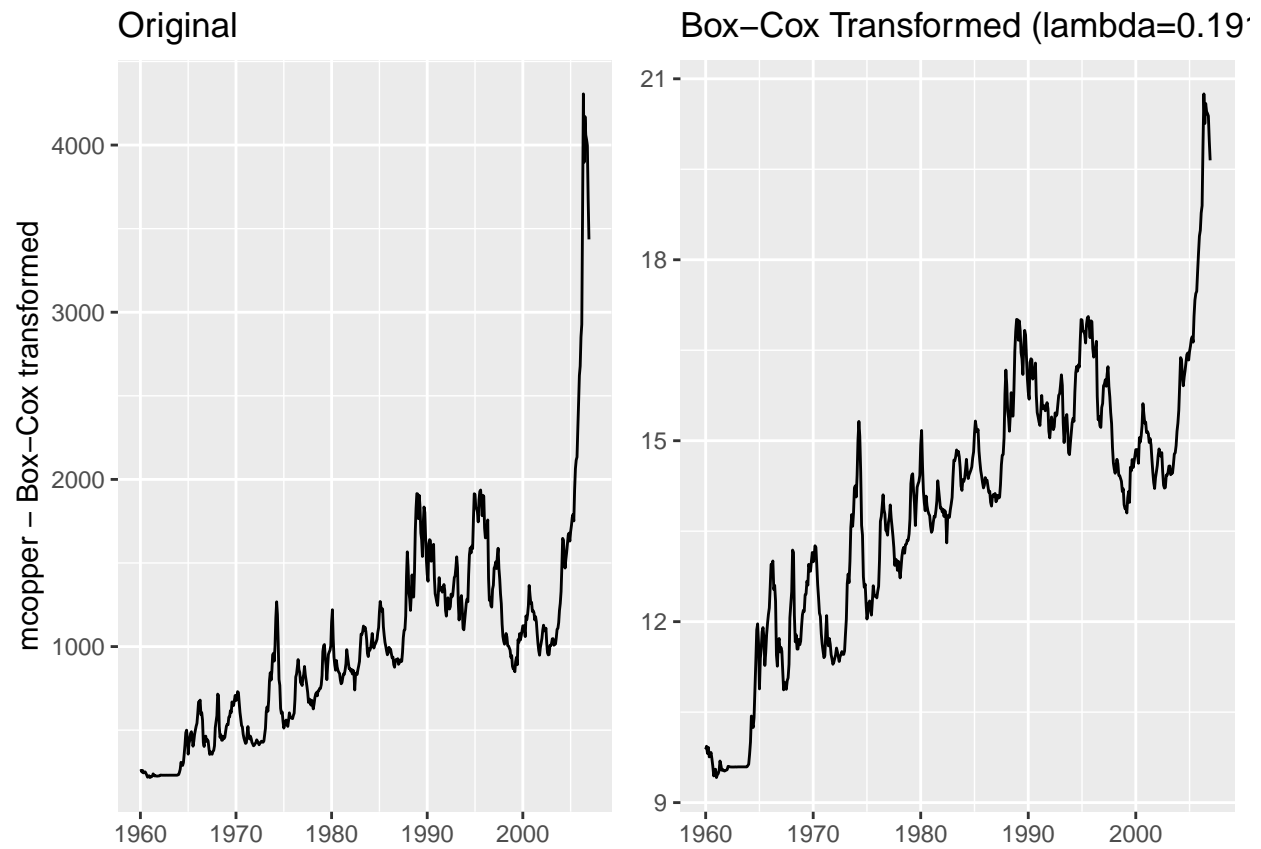
usgdp

```
boxcox_trans_loop(usgdp, "usgdp - Box-Cox transformed")
```



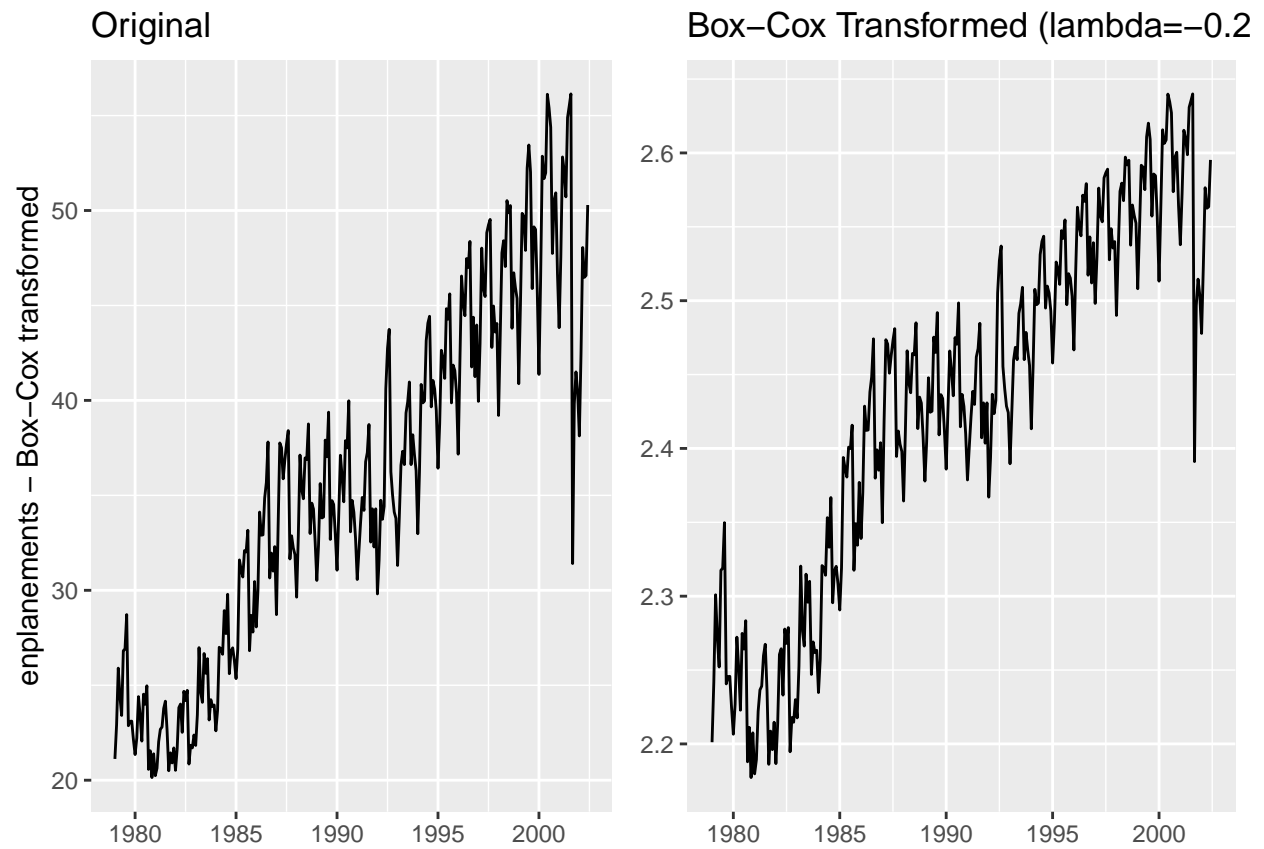
mcopper

```
boxcox_trans_loop(mcopper, "mcopper - Box-Cox transformed")
```



enplanements

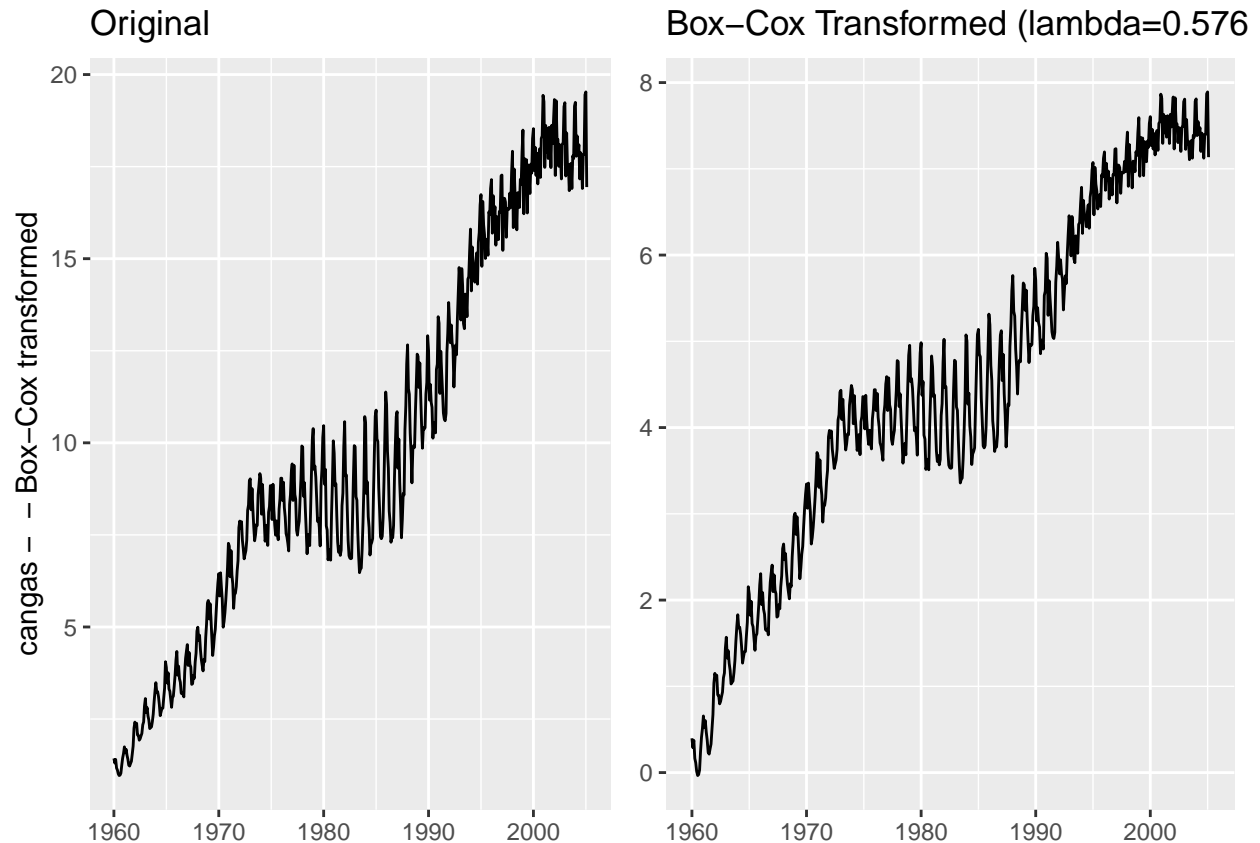
```
boxcox_trans_loop(enplanements, "enplanements - Box-Cox transformed")
```



Exercise 3.2

Why is a Box-Cox transformation unhelpful for the `cangas` data?

```
boxcox_trans_loop(cangas, "cangas - - Box-Cox transformed")
```



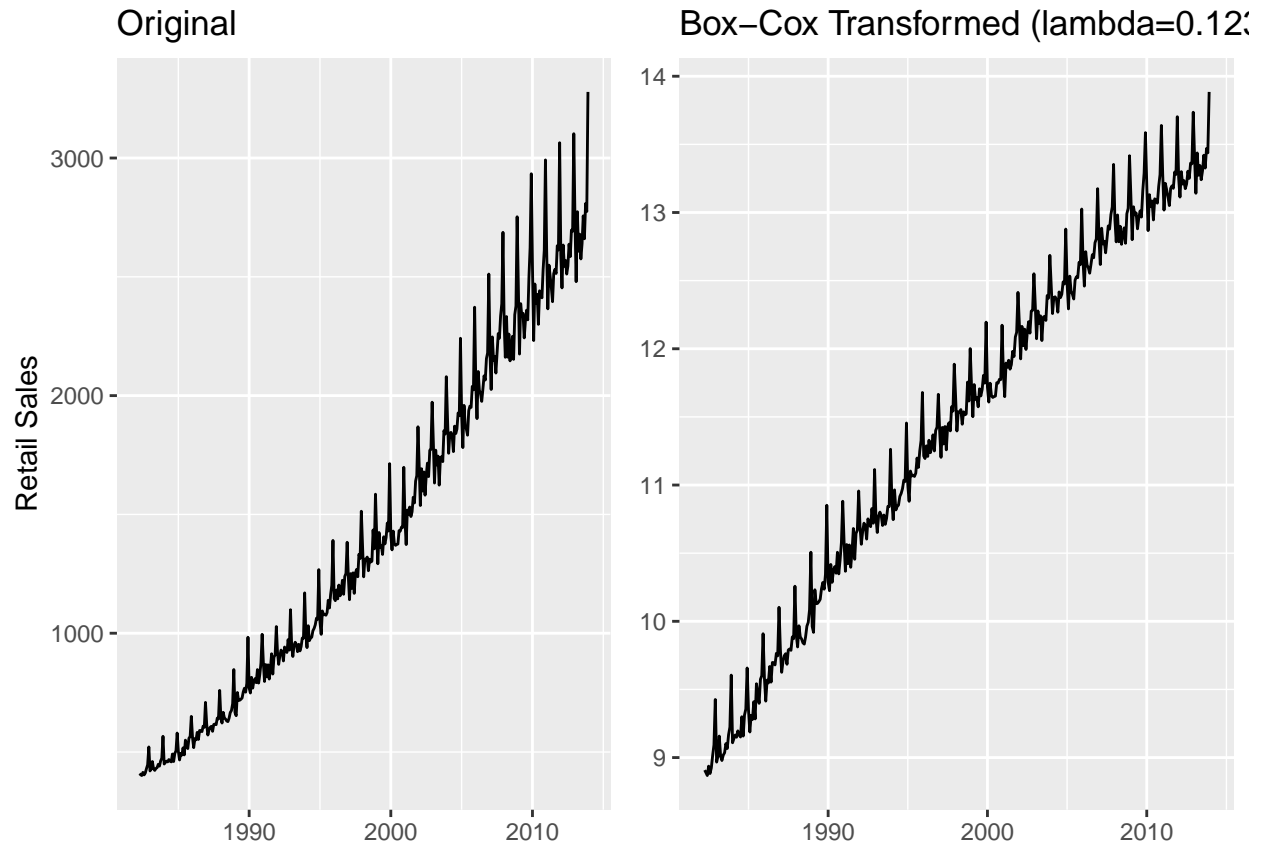
In above plot, we see that seasonal variability remains largely same throughout the years, where variability increased after 1975 and reduced after 1990.

So to stabilise the variance, the transformation will need to “stretch” the seasonal variability before 1975; “shrink” the variability between 1975 and 1985; and then again “stretch” the variability after 1990.

Exercise 3.3

What Box-Cox transformation would you select for your `retail` data?

```
retaildata <- read_excel("retail.xlsx", skip = 1)
myts <- ts(retaildata[, "A3349398A"], frequency = 12, start = c(1982, 4))
# myts <- ts(retaildata[, "A3349709X"], frequency = 12, start = c(1982, 4))
# myts <- ts(retaildata[, "A3349413L"], frequency = 12, start = c(1982, 4))
# myts <- ts(retaildata[, "A3349335T"], frequency = 12, start = c(1982, 4))
# myts <- ts(retaildata[, "A3349627V"], frequency = 12, start = c(1982, 4))
# myts <- ts(retaildata[, "A3349338X"], frequency = 12, start = c(1982, 4))
# myts <- ts(retaildata[, "A3349398A"], frequency = 12, start = c(1982, 4))
boxcox_trans_loop(myts, "Retail Sales")
```



In the original data, variance increased from 1990 to 2010+. But, in Box-cox transformed data, variance seems to have become uniform. The value for Lambda is .1232. So, this transformation appears to be effective.

Exercise 3.8

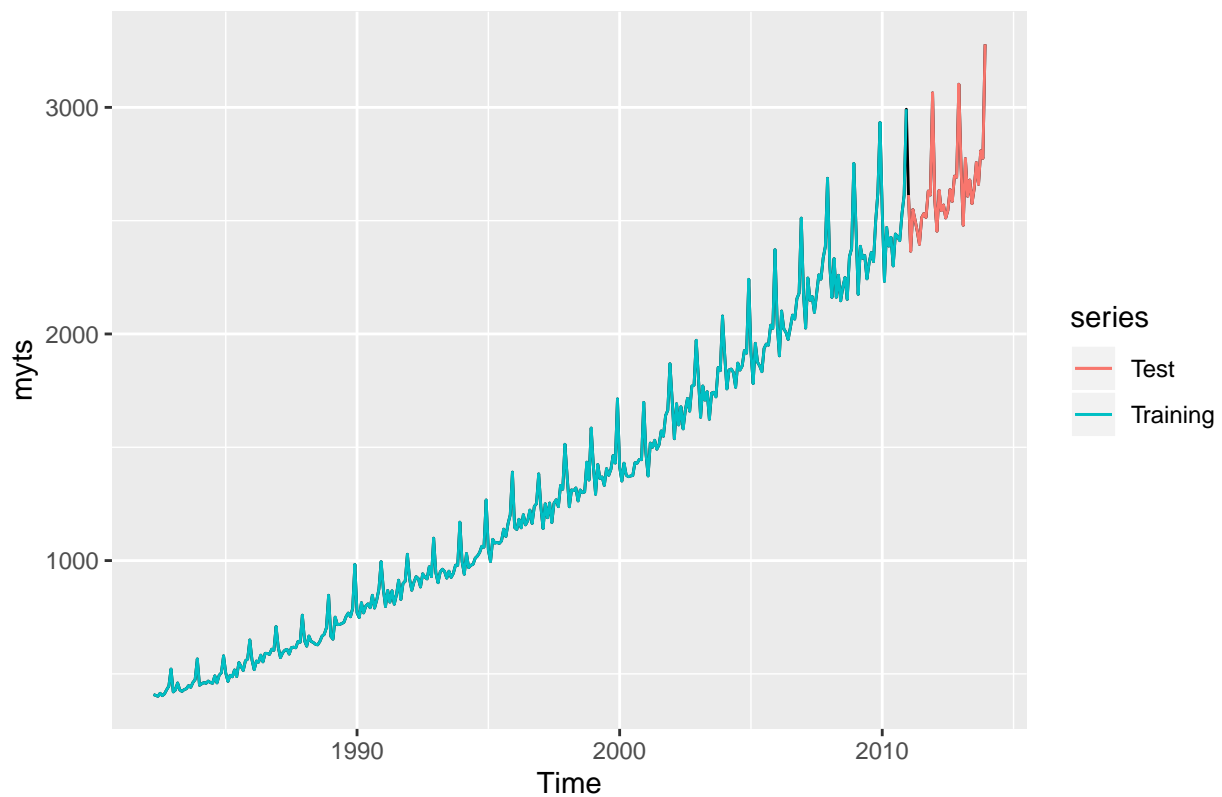
For your `retail` time series:

- a) Split the data into two parts using

```
myts.train <- window(myts, end = c(2010, 12))
myts.test <- window(myts, start = 2011)
```

- b) Check that your data have been split appropriately by producing the following plot.

```
autoplot(myts) +
  autolayer(myts.train, series = "Training") +
  autolayer(myts.test, series = "Test")
```



c) Calculate forecasts using `snaive` applied to `myts.train`.

```
fc <- snaive(myts.train)
```

d) Compare the accuracy of your forecasts against the actual values stored in `myts.test`.

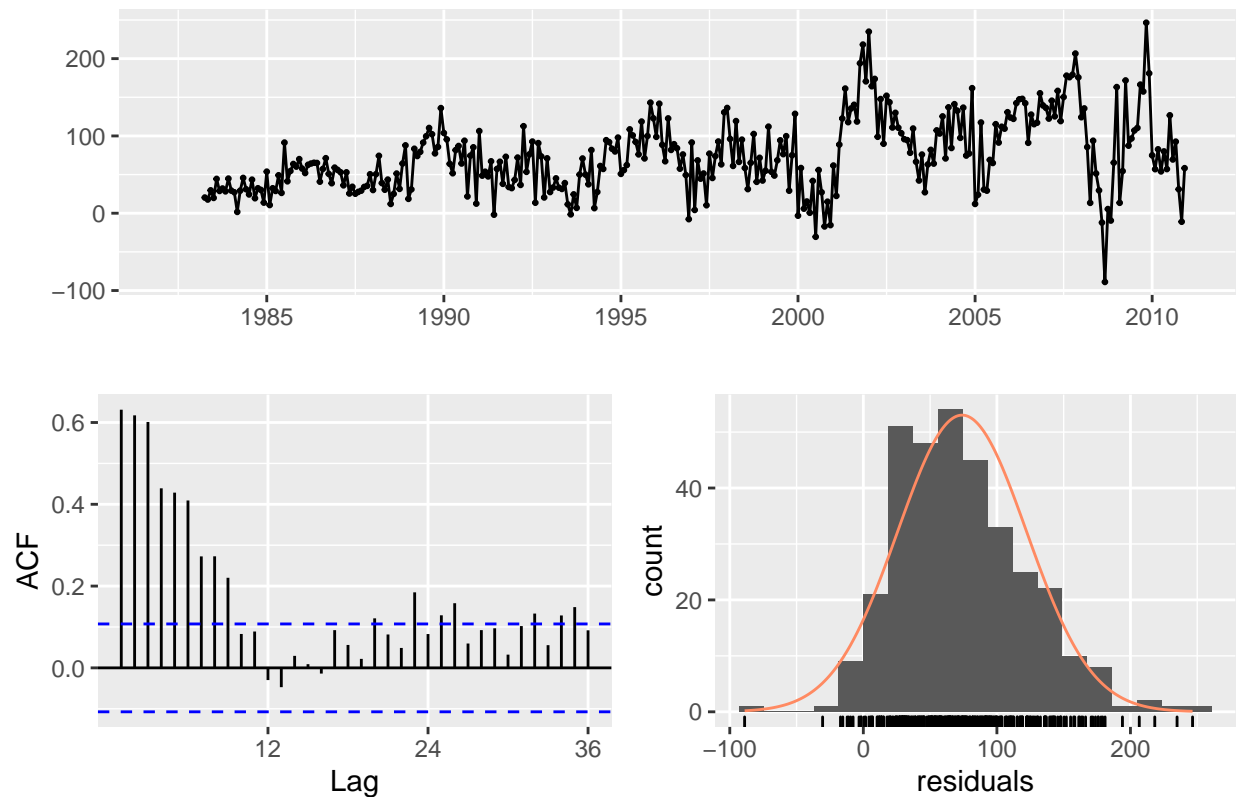
```
accuracy(fc, myts.test)
```

```
##           ME      RMSE      MAE      MPE      MAPE      MASE
## Training set  73.94114  88.31208  75.13514  6.068915  6.134838  1.000000
## Test set     115.00000 127.92727 115.00000  4.459712  4.459712  1.530576
##           ACF1 Theil's U
## Training set  0.6312891    NA
## Test set     0.2653013 0.7267171
```

e) Check the residuals.

```
checkresiduals(fc)
```


Residuals from Seasonal naive method



```
##
##  Ljung-Box test
##
## data:  Residuals from Seasonal naive method
## Q* = 671.41, df = 24, p-value < 2.2e-16
##
## Model df: 0.   Total lags used: 24
```

Do the residuals appear to be uncorrelated and normally distributed?

Yes, the residuals seem to be normally distributed but with a bit of a positive skew. The residuals don't appear to be uncorrelated. The Ljung-Box test shows a p value of that less than 0.05, which suggests incomplete information, and therefore the seasonal naive model is not reliable model.

f) How sensitive are the accuracy measures to the training/test split?

The accuracy measures are quite sensitive to the training/test split. The values are significantly different between the two. Probably the model doesn't generalize well.